UNIVERSITY OF CALIFORNIA – LOS ANGELES

all caps

(6) Bounds for Iterates, Inverses and Spectral Variation of Non-Normal Matrices

(7) NA
(8) NA
(9) NA
(13) 97 p.
(13) NA
(14) NA
(12) NA
(13) NA
(14) NA

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Mathematics

(16) by

(5) Edward Arthur Sallin

Final Examination for the Degree Doctor of Philosophy
Thursday, May 23, 1963, 9:30 A.M.
Room 6118, Mathematical Sciences

Committee in charge:

Professor Peter K. Henrici, Chairman
Professor Edwin F. Beckenbach, Co-Chairman
Professor Alfred Horn
Associate Professor Michel A. Melkanoff
Associate Professor Robert H. Sorgenfrey
Associate Professor Donald H. Stork

(11) June, 1963

(15) Contract NONR 2336(16), Proj. NR-044-144/11-29-61

(1) #9.10

The dissertation of Edward Arthur Sallin is approved, and it is acceptable in quality and form for publication on microfilm:

_____

_____

_____

_____

_____

Co-Chairman

_____

Committee Chairman

University of California, Los Angeles

May, 1963

TABLE OF CONTENTS

ACKNOWLEDGEMENT

The author wishes to express his deep gratitude to Professor Peter K. Henrici for his personal interest and guidance in the preparation of this dissertation.

Thanks are also due to the members of the staff of Numerical Analysis Research and of the Department of Mathematics for their help while the author was preparing for this work.

Special thanks are due to Mrs. Elaine Barth for her suggestions and excellent work preparing the manuscript.

This dissertation is lovingly dedicated to my best friend and wife, Norma K. Sallin.

The preparation of this paper was sponsored by the Office of Naval Research, U. S. Navy to whom I am indebted. Reproduction in whole or in part is permitted for any purpose of the United States Government.

VITA

April 29, 1937 – Born – Pittsburgh, Pennsylvania

1958 – A.B., University of California, Los Angeles

1958–1960 – Research Assistant, Numerical Analysis Research, Department of Mathematics, University of California, Los Angeles

1960 – M.A., University of California, Los Angeles

1960–1962 – Graduate Research Mathematician, Numerical Analysis Research, Department of Mathematics, University of California, Los Angeles

1962 – Teaching Assistant, Department of Mathematics, University of California, Los Angeles, California

1962–1963 – Member of the Technical Staff, Aerospace Corp., El Segundo, California

FIELDS OF STUDY

Major Field: Mathematics

Studies in Numerical Analysis. Professor Peter K. Henrici

Studies in Real Analysis. Professors John W. Green and Angus E. Taylor

Studies in Complex Analysis. Professors Robert H. Sorgenfrey and Stefan E. Warschawski

Studies in Functional Analysis. Professor Angus E. Taylor

Studies in Algebra. Professor Lowell J. Paige

ABSTRACT OF THE DISSERTATION

Bounds for Iterates, Inverses and Spectral Variation
of Non-Normal Matrices
by
Edward Arthur Sallin

University of California, Los Angeles, 1963
Professor Peter K. Henrici, Chairman

For an arbitrary multiplicative matrix norm ν and arbitrary non-singular matrix N, we define the condition number $C_\nu(N)$ of N with respect to ν as $\nu(N)\,\nu(N^{-1})$. A square matrix is said to be quasi-diagonal if it is a symetrically partitioned triangular matrix which is diagonal when considered as a partitioned matrix.

The following problems of computational linear algebra are considered in this paper:

(I) Given an arbitrary square matrix A, to explicitly construct and determine a bound for a condition number of a matrix N such that $Q = N^{-1}AN$ is quasi-diagonal.

(II) To estimate the norms of A^n, $n = 1,2,\ldots$ in terms of the eigenvalues of A and $C_\nu(N)$.

(III) To estimate the error $\tilde{x} - A^{-1}b$ of an approximate solution $\tilde{x}$ of the equation $Ax = b$ in terms of the residual $r = A\tilde{x} - b$, the eigenvalues of A and $C_\nu(N)$. To estimate the error $\tilde{X} - A^{-1}$ of an approximate inverse $\tilde{X}$ of A in terms of the residual $A\tilde{X} - I$, the eigenvalues of A and $C_\nu(N)$.

(IV) To estimate the distance of the spectrum of an arbitrary matrix B from the spectrum of A in terms of a norm of $B - A$, eigenvalues of A and $C_\nu(N)$.

Solutions to problems (II), (III) and (IV) are classical if A is normal, i.e., $AA^* = A^* A$. Solutions have been constructed for non-normal A, but with less satisfactory results. A partitioned Schur form of a given matrix A will be called an ordered Schur form if (i) the eigenvalues are lexicographically ordered by blocks on the diagonal, and (ii) equal eigenvalues belong to the same block. Let A be an ordered Schur form of B and let N be chosen such that $Q = N^{-1} AN$ is quasi-diagonal with $Q = \operatorname{diag}(Q_{11}, Q_{22}, \ldots, Q_{kk})$, where $Q_{ii} = A_{ii}$ is of order n_i. Writing $Q_{ii} = D_i + L_i$ where D_i is the diagonal part of Q_{ii} and setting $\ell_i = \sigma(L_i)$, $\Delta_i = \lambda_{Q_{ii}}$ where σ is the spectral norm and λ_A is the spectral radius of A, we can conclude:

THEOREM. If $\lambda_B > 0$,

$$\sigma(B^r) \leq \min\left[C_\sigma(N) \max_i \left\{\Delta_i^r + \binom{r}{1}\Delta_i^{r-1}\ell_i + \cdots + \binom{r}{n_i - 1}\Delta_i^{r-n_i+1}\ell_i^{n_i-1}\right\}\right].$$

If $\lambda_B = 0$,

$$\sigma(B^r) \leq \min\left[C_\sigma(N) \max_i \ell_i^r\right], \qquad r = 1, 2, \ldots, M-1$$

$$\sigma(B^r) = 0, \qquad r \geq M$$

where $M = \max n_i$ and where the minimum is taken over all ordered Schur forms.

If $f^n(x)$ is defined for all $x \geq 0$ by $f^n(x) = x + x^2 + \cdots + x^n$ we have:

THEOREM. If B is non-singular and non-normal and if

$$\xi_1 = \Delta_1^{-1} \ell_1$$

then

$$\sigma(B^{-1}) \leq \min\left[C_\sigma(N) \max_i \left\{ \frac{f^{n_i}(\xi_i)}{\xi_i} \Delta_i^{-1} \right\}\right]$$

where the minimum is taken over all ordered Schur forms.

Let the function $g^n = g^n(y)$ be defined for all $y \geq 0$ as the (unique) non-negative solution of the equation $g + g^2 + \cdots + g^n = y$. Then if M is an arbitrary matrix with eigenvalues λ_i and B has eigenvalues μ_i, the quantity

$$s = s_M(B) = \max_i \left\{\min_j |\mu_i - \lambda_j|\right\}$$

is called the spectral variation of B with respect to M.

THEOREM. For non-normal M with $M - B \neq 0$ we have for any norm ν dominating σ:

$$s_M(B) \leq \min \left\{\left[\max_i \frac{y_i}{g^{n_i}(y_i)}\right] C_\nu(N)\, \nu(U^*BU - M)\right\}$$

where

$$y_i = \frac{\nu(L_i)}{C_\nu(N)\, \nu(U^*BU - M)}$$

and the minimum is taken with respect to all U occurring in an ordered Schur form of M.

INTRODUCTION

Unless otherwise stated, all matrices in this paper shall be assumed to be rectangular $m \times n$ with complex elements and shall be denoted by $A = (a_{ij})$, a vector shall mean a column vector with n complex elements.

A matrix norm is a real valued function ν defined on the space of square matrices and satisfying the following relations for arbitrary matrices A and B and arbitrary complex scalars c:

(a) $\nu(A) \geq 0$; $\nu(A) = 0$ if and only if $A = 0$

(b) $\nu(cA) = |c| \, \nu(A)$

(c) $\nu(A + B) \leq \nu(A) + \nu(B)$.

If in addition,

(d) $\nu(AB) \leq \nu(A) \, \nu(B)$

the norm is called multiplicative. We shall be concerned primarily with such norms.

A vector norm is a real-valued function defined on the space of vectors and satisfying relations analogous to (a), (b) and (c) above.

By the spectrum of a matrix A we mean the totality of its eigenvalues, considered as a point set in the complex plane. The largest of the moduli of the eigenvalues of A is called the spectral radius of A and is denoted by λ_A. For any invertible

matrix S and multiplicative norm ν the condition number, $C_\nu(S)$, of S with respect to the norm ν is defined by

$$C_\nu(S) = \nu(S)\,\nu(S^{-1}). \qquad \text{Bauer [2]}$$

A matrix is said to be partitioned (a partitioned or block matrix) if it has been divided into smaller arrays by horizontal and vertical lines and each of the resulting submatrices have been represented by a single element. If an $m \times n$ matrix A is partitioned it shall be denoted by $A = (A_{ij})$; the i,j element of A being the submatrix A_{ij} of A, or order $m_i \times n_j$ where $\sum m_i = m$, $\sum n_j = n$. A partitioning of a matrix A by an equal number of horizontal and vertical lines in such a manner that each of the resulting diagonal entries, A_{ii}, are square matrices is called a symmetric partition.

A symmetrically partitioned triangular matrix A will be called sp-triangular. Thus

$$A_{ij} = 0 \quad \text{for} \quad i < j, \quad \text{and}$$

$$A_{ii} \quad \text{are (lower) triangular matrices.}$$

A partitioned matrix A is said to be quasi-diagonal if it is sp-triangular and if $A_{ij} = 0$ for $i \neq j$.

The following problems of computational linear algebra will be considered in this paper:

(1) Given an arbitrary sp-triangular matrix, A, to construct a matrix N such that $N^{-1}AN$ is quasi-diagonal with

predetermined diagonal entries and to determine a bound for a condition number of N.

(ii) To estimate the norms of the matrices A^n, $n = 1,2,\ldots$, in terms of the eigenvalues of A and a condition number of N, above.

(iii) To estimate the error $\tilde{x} - A^{-1}b$ of an approximate solution $\tilde{x}$ of the equation $Ax = b$ in terms of the residual $r = A\tilde{x} - b$, the eigenvalues of A and a condition number of N.

(iv) To estimate the distance of the spectrum of a matrix B from the spectrum of A in terms of a norm of $B - A$, the eigenvalues of A and a condition number of N.

Solutions to problems (ii), (iii) and (iv) are classical if A is normal, i.e., $AA^* = A^*A$. Solutions have been constructed for non-normal A, but with less satisfactory results. Some of the bounds given depend on a knowledge of a matrix S in the representation $A = SJS^{-1}$, when J is the Jordan canonical form. Other bounds do not approach the classical bounds if A approaches a normal matrix. The bounds given in the present paper, while depending on the eigenvalues of A and their multiplicities, do not require a knowledge of the Jordan canonical form. Furthermore, our estimates approach the classical estimates for A normal. Our insistence on not using the Jordan form is motivated partly by reasons of computational convenience, and partly by the fact that the Jordan form is a discontinuous function on the space of matrices and is therefore ill suited for purposes of computation (see [8]

The enclosed Technical Report

was prepared at

Numerical Analysis Research
Department of Mathematics
The University of California
Los Angeles 24, California

under the sponsorship of

one or more of the following:

OFFICE OF NAVAL RESEARCH

Project No. NR-044-144/11-29-61
Contract Nonr-233(76)

U. S. ARMY RESEARCH OFFICE (DURHAM)

Project No. 1210-M
Grant DA-ARO(D)-31-124-G77

for related remarks).

We shall develop a canonical form for arbitrary (non-normal) matrices [problem (i) above] which is continuous in nature and for which we can explicitly demonstrate a transforming matrix N and its condition number. Such a canonical form is developed in 2.1 and 3.1. Representations for N are given in 2.21 and 3.2 and estimates for a condition number are given in sections 2.22 and 3.2.

These results are then applied to problems (ii), (iii) and (iv) in Chapters 4, and 5.

Certain estimates based upon a measure of non-normality of a matrix have been derived by Wielandt in [29] and Henrici in [14]. Wielandt's measure is applicable only to matrices which are similar to a diagonal matrix. Henrici removes this restriction but gets estimates in terms of λ_A, consequently making no use of eigenvalues of smaller modulus

CHAPTER 1

Preliminaries on Norms

It will be necessary to consider norms defined for rectangular matrices. That this can be done is shown by the following lemma.

LEMMA 1. *Given a family* F *of rectangular matrices of bounded row and column dimension, say* k, *and an arbitrary multiplicative norm* ν *defined for square matrices, then exists a family of norms* $\nu_q : q \geq k$ *which are multiplicative on* F.

Proof. Let A,B,C be members of F where A and B are of order $r_1 \times s_1$ and C is of order $r_2 \times s_2$. Let q be any integer such that $q \geq k$. In particular then $q \geq r_1, s_1, r_2, s_2$. Define A_q to be the $q \times q$ matrix formed from A by the addition of $q - r_1$ rows of zeros and $q - s_1$ columns of zeros:

$$A_q = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r_1 \\ q-r_1 \end{matrix}, \qquad \begin{matrix} s_1 & q-s_1 \end{matrix}$$

Define $\nu_q(A) = \nu(A_q)$.

With this definition ν_q has the properties of a multiplicative norm, for

(i) $\nu_q(A) = 0$ implies $\nu(A_q) = 0$ which in turn means that A_q and consequently A are null matrices. $\nu_q(A) \geq 0$ since $\nu(A_q) = \nu_q(A)$.

(ii) $\nu_q(cA) = \nu\left[(cA)_q\right] = \nu(cA_q) = |c|\ \nu(A_q) = |c|\ \nu_q(A)$.

(iii) $\nu_q(A + B) = \nu\left[(A + B)_q\right] = \nu\left[A_q + B_q\right]$

$$\leq \nu(A_q) + \nu(B_q) = \nu_q(A) + \nu_q(B).$$

Whenever the product AC is defined, i.e., when $s_1 = r_2$ we have

(iv) $$(AC)_q = \begin{pmatrix} AC & 0 \\ 0 & 0 \end{pmatrix}^{r_1}_{q-r_1} = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}^{r_1}_{q-r_1} \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}^{s_1}_{q-s_1}$$

$$= A_q C_q, \text{ and}$$

$$\nu_q(AC) = \nu\left[(AC)_q\right] = \nu(A_q C_q)$$

$$\leq \nu(A_q)\ \nu(C_q) = \nu_q(A)\ \nu_q(C).$$

We shall have occasion to deal with functions defined on scalar matrices whose elements are themselves norms of elements of some fixed partitioned matrix. By restricting the class of norms employed we can guarantee that these functions will be norms of the original matrix. The principal result is given by Lemma 3.

If A and B are matrices of the same row and column dimensions, $A \leq B$ shall mean $a_{ij} \leq b_{ij}$ $(i = 1,2,\ldots,m;$ $j = 1,2,\ldots,n)$. Given any matrix $A = (a_{ij})$, $|A|$ is the matrix whose general element is $|a_{ij}|$. A norm ν is called <u>monotone</u> if for A and B of the same dimensions, $|A| \leq B$ implies $\nu(A) \leq \nu(B)$

A sufficient condition for a norm to be monotone is given by the following lemma.

LEMMA 2. *Let ν be any norm such that $\nu(|A|) = \nu(A)$. Then ν is monotone.*

Proof. Let A and B be given matrices of order $m \times n$ and let $|A| \leq B$. Thus

$$|a_{ij}| \leq b_{ij} \qquad (i = 1,2,\ldots,m;\ j = 1,2,\ldots,n). \tag{1.1}$$

Since, by hypothesis, ν depends only upon the magnitude of each element of A we can assume that all $a_{ij} \geq 0$. To show that $\nu(A) \leq \nu(B)$ it is sufficient to consider the case where only one equality in (1-1) fails to hold, say $a_{rs} < b_{rs}$.

By postulate (b) of the definition of a norm we can assume that $b_{rs} = 1$. To simplify the writing we can assume further that we have $r = s = 1$ and hence $0 \leq a_{11} < b_{11} = 1$ and have to show that $\nu(a_{11}, b_{12},\ldots,b_{mn}) \leq \nu(1, b_{12},\ldots,b_{mn})$. But this follows immediately if we use the decomposition

$$\begin{pmatrix} a_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} = \frac{1 + a_{11}}{2} \begin{pmatrix} 1 & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

$$+ \frac{1 - a_{11}}{2} \begin{pmatrix} -1 & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

and apply the triangle inequality and postulate (b) of a norm, since we then have

$$\begin{aligned}
&\nu(a_{11},b_{12},\ldots,b_{mn}) \\
&\quad \leq \frac{1+a_{11}}{2}\,\nu(1,b_{12},\ldots,b_{mn}) + \frac{1-a_{11}}{2}\,\nu(-1,b_{12},\ldots,b_{mn}) \\
&\quad = \frac{1+a_{11}}{2}\,\nu(1,b_{12},\ldots,b_{mn}) + \frac{1-a_{11}}{2}\,\nu(1,b_{12},\ldots,b_{mn}) \\
&\quad = \nu(1,b_{12},\ldots,b_{mn}).
\end{aligned}$$

This generalizes Ostrowski's [20] concept of coordinatewise symmetric gauge functions.

The use we shall make of the concept of monotone norms is given in the following lemma.

LEMMA 3. Let $A = (A_{ij})$ be a partitioned matrix. Let ρ be an arbitrary multiplicative norm and $\tilde{A}$ the scalar matrix whose general element is $\rho(A_{ij})$.

Then if ν is a monotone multiplicative norm, the function N, defined by $N(A) = \nu(\tilde{A})$ is a multiplicative norm.

Proof. We must verify the four postulates for a multiplicative norm. Namely,

(a) $N(A) = 0$ implies $\nu(\tilde{A})$ and hence $\tilde{A} = 0$. Then $\rho(A_{ij}) = 0$, $A_{ij} = 0$ and finally $A = 0$. $N(A) \geq 0$ since ν is a norm.

$$(b)\quad N(cA) = \mathcal{V}(c\tilde{A}) = \mathcal{V}\left[\rho(cA_{ij})\right] = \mathcal{V}\left[|c|\rho(A_{ij})\right]$$

$$= \mathcal{V}(|c|\tilde{A}) = |c|\,\mathcal{V}(\tilde{A}) = |c|\ N(A)$$

$$(c)\quad N(A+B) = \mathcal{V}(\widetilde{A+B})$$

$$(1\text{-}2)\qquad = \mathcal{V}\left[\rho(A_{ij}+B_{ij})\right] \le \mathcal{V}\left[\rho(A_{ij})+\rho(B_{ij})\right]$$

$$= \mathcal{V}(\tilde{A}+\tilde{B}) \le \mathcal{V}(\tilde{A}) + \mathcal{V}(\tilde{B}) = N(A) + N(B)$$

$$(d)\quad N(AB) = \mathcal{V}(\tilde{A}\tilde{B})$$

$$(1\text{-}3)\qquad = \mathcal{V}\left[\rho\Big(\sum_k A_{ik}B_{kj}\Big)\right] \le \mathcal{V}\left[\sum_k \rho(A_{ik})\,\rho(B_{kj})\right]$$

$$= \mathcal{V}(\tilde{A}\tilde{B}) \le \mathcal{V}(\tilde{A})\,\mathcal{V}(\tilde{B}) = N(A)N(B)$$

(1.2) and (1.3) hold since $\mathcal{V}$ is monotone.

The following are some of the most common norms of matrices $A = (a_{ij})$. (See [16], [21]).

$$\alpha(A) = \sum_{i,j} |a_{ij}|$$

$$\sigma(A) = \max_{x\neq 0}\left[\frac{x^*A^*Ax}{x^*x}\right]^{1/2} \qquad \text{(Spectral norm)}$$

$$\rho(A) = \max_i \sum_j |a_{ij}|$$

$$\gamma(A) = \max_j \sum_i |a_{ij}|$$

$$\epsilon(A) = \left[\sum_{i,j} |a_{ij}|^2\right]^{1/2} \qquad \text{(Euclidean norm)}$$

The last three are obvious examples of monotone norms.

If ϕ is a vector norm, then the function ν_ϕ defined by

$$\nu_\phi = \sup_{x \neq 0} \frac{\phi(Ax)}{\phi(x)}$$

always defines a matrix norm. Matrix norms defined in this way are called lub norms in [3]. The norms σ, ρ and γ defined above can be derived in this manner from suitable vector norms [24]. On the other hand, some matrix norms, such as ϵ, cannot be thus derived.

We shall use the following definitions:

A matrix norm ν is called compatible with a vector norm μ, if $\mu(Ax) \leq \nu(A)\mu(x)$ for all matrices A and vectors x. A lub norm is always compatible with the vector norm defining it.

A matrix norm ν will be called unitarily invariant, if $\nu(U^*AU) = \nu(A)$ for all A and all unitary U. The norms σ and ϵ are unitarily invariant, while σ and γ are not.

A lub norm is called axis-oriented [3] if $\nu(D) = \max_{1 \leq i \leq n} |d_{ii}| = \lambda_D$ for any diagonal matrix $D = (d_{ij})$. The lub norms σ, ρ, γ are axis-oriented.

A norm ν is said to majorize another norm μ if $\nu(A) \geq \mu(A)$ for all A. The ϵ norm majorizes σ.

We shall require the following consequences of the defining properties of a norm. [See [21] for proofs].

LEMMA 4. If λ_A denotes the spectral radius of A, then $\nu(A) \geq \lambda_A$ for any matrix norm ν.

LEMMA 5. If μ and ν are any two matrix norms, then there exists a constant $P_{\mu\nu}$, depending only on these two norms, such that

$$\mu(A) \leq P_{\mu\nu}\, \nu(A),$$

for all matrices A.

Values of $P_{\mu\nu}$ for special norms are given in [27].

We have finally

LEMMA 6. Let D be the quasi-diagonal matrix $D = dg(D_1, D_2, \ldots, D_k)$. Then $\sigma(D) = \max_i \sigma(D_i)$.

For the proof we note first that $\sigma^2(A) = \lambda_{A^*A}$ for all A and in particular $\sigma^2(D) = \lambda_{D^*D}$. But $D^*D = dg(D_1^*D_1, D_2^*D_2, \ldots, D_k^*D_k)$ and the eigenvalues of D^*D are the union of those of $D_i^*D_i$. Thus $\lambda_{D^*D} = \max_i \lambda_{D_i^*D_i}$ and $\sigma^2(D) = \max_i \sigma^2(D_i)$.

CHAPTER 2

Introduction. This chapter is divided into two sections. The first is concerned with the problem of establishing the existence of a matrix N such that $N^{-1}BN$ is quasi-diagonal for an arbitrary matrix B. The results of this section are used to give an explicit representation for N, above, and to estimate its condition number.

Section 2.1. We may restrict our attention to any one of the triangular forms A of a given matrix B since every matrix is unitarily similar to a triangular matrix, and since we shall ultimately make use only of unitarily invariant norms. It is true further that the ordering of the diagonal elements of A (the eigenvalues of A) may be specified arbitrarily. It is of importance to subsequent estimates we shall make that the specification of the ordering of eigenvalues does not uniquely determine the triangular form.

We assume then that

$$\lambda_1 \preceq \lambda_2 \preceq \cdots \preceq \lambda_n$$

where $\lambda_i = a_{ii}$, $i = 1,2,\ldots,n$ are the eigenvalues of A, and where $\lambda_i \prec \lambda_j$ means that either

(a) $\operatorname{Re} \lambda_i < \operatorname{Re} \lambda_j$ or

(b) $\operatorname{Re} \lambda_i = \operatorname{Re} \lambda_j$ and $\operatorname{Im} \lambda_i < \operatorname{Im} \lambda_j$.

This is the so-called lexicographic, or dictionary, ordering of the complex plane.

If A has eigenvalues $\lambda_1,\lambda_2,\ldots,\lambda_n$ we define the (disjoint) sets $S_1,S_2,\ldots,S_k$ of order $n_1,n_2,\ldots,n_k$ respectively, such that $\sum_{i=1}^{k} n_i = n$, and

(a) $\lambda_i \in S_k$, $\lambda_j = \lambda_i$ implies $\lambda_j \in S_k$ and

(b) $\lambda_i \in S_k$, $\lambda_j \in S_p$, $k < p$ implies $\lambda_i \prec \lambda_j$.

These sets S_k uniquely determine a symmetric partition of A if we specify that the diagonal terms of each diagonal submatrix resulting from the partitioning belong to one and only one S_j.

Any matrix which satisfies the above triangularity, ordering and partition conditions shall be said to be in an ordered Schur form or to be an ordered Schur matrix.

Let then $A = (A_{ij})$, $(i,j = 1,2,\ldots,k)$ be any ordered Schur matrix, which we assume is of order k, and where the A_{ij} are matrices of order $n_i \times n_j$ with $\sum_{i=1}^{k} n_i = n$.

DEFINITION. E_i denotes the row vector,

$$(E_{i1},E_{i2},\ldots,E_{ik})$$

where E_{is} is of order $n_i \times n_s$ and

$$E_{is} = \begin{cases} 0 & \text{if } s \neq i \\ I_{n_i} & \text{if } s = i \end{cases}$$

Here I_{n_i} denotes the $n_i \times n_i$ identity matrix and I (with no subscripts) is the $n \times n$ identity matrix partitioned as A above.

DEFINITION. E_i^T is the column vector

$$\begin{bmatrix} E_{i1}^T \\ E_{i2}^T \\ \vdots \\ E_{ik}^T \end{bmatrix}$$

where the E_{is}^T are the transposes of the E_{is}. Then

$$E_i E_j^T = E_j E_i^T = \begin{cases} 0 & , \quad j \neq i \\ I_{n_i} & , \quad j = i \end{cases} \tag{2.1-1}$$

It then follows that

$$A_{ij} = E_i A E_j^T \tag{2.1-2}$$

The matrix $M = E_i^T N_{ij} E_j$, where N_{ij} is an arbitrary $n_i \times n_j$ matrix, is such that

$$M_{rs} = \begin{cases} N_{ij} & \text{for } r = i, \ s = j \\ 0 & \text{otherwise} \end{cases} \tag{2.1-3}$$

DEFINITION. The partitioned matrices E_{ij} defined by

$$E_{ij} \equiv E_{ij}[E_i, E_j ; N_{ij}] = I + E_i^T N_{ij} E_j$$

where the N_{ij} are arbitrary $n_i \times n_j$ matrices for $i > j$ and null matrices otherwise are called <u>elementary block matrices</u>. That

is

$$(E_{ij})_{rs} = \begin{cases} I_{n_i} & r = s = i \\ N_{ij} & r = i,\ s = j \\ 0 & \text{otherwise} \end{cases}$$

Using (2.1–1), (2.1–2) and (2.1–4), the following properties of elementary block matrices may be verified:

$$E_{ij}^{-1}\left[E_i, E_j\ ;\ N_{ij}\right] = E_{ij}\left[E_i, E_j\ ;\ -N_{ij}\right]$$

$$E_{ij}\,E_{kj} = E_{ij} + E_{kj} - I.$$

Letting

$$N_j = \prod_{i=1}^{k} E_{ij} = \prod_{i>j} E_{ij} \qquad (j = 1,2,\ldots,k-1) \tag{2.1–5}$$

we have

$$N_j = I + \Big(\sum_{i>j} E_i^T N_{ij}\Big)E_j.$$

That is

$$(N_j)_{rs} = \begin{cases} I_{n_i} & r = s = i \\ N_{rj} & r > s,\ s = j \\ 0 & \text{otherwise} \end{cases} \qquad i = 1,2,\ldots,k$$

Furthermore,

$$N_j^{-1} = I - \Big(\sum_{i>j} E_i^T N_{ij}\Big)E_j.$$

Letting

(2.1-6) $$N = \prod_{j=1}^{k-1} N_j; \qquad \text{we have}$$

$$(N)_{rs} = \begin{cases} I_{n_i} & r = s = i \qquad (i = 1,\ldots,k) \\ N_{rs} & r > s \\ 0 & r < s \end{cases}$$

$$N^{-1} = \prod_{j=0}^{k-1} N_{k-j}^{-1}.$$

Letting $\tilde{N} = (\tilde{N}_{ij})$, where

$$\tilde{N}_{ij} = \begin{cases} N_{ij} & i > j \\ 0 & i \leq j \end{cases} \qquad \text{we have}$$

since $\tilde{N}^r = 0$ for $r \geq k$

$$N^{-1} = (I + \tilde{N})^{-1} = I - \tilde{N} + \tilde{N}^2 - \cdots (-1)^{k-1} \tilde{N}^{k-1}.$$

DEFINITION. The function F_{ij} defined by

$$F_{ij}(A) = E_{ij}^{-1} A E_{ij}, \qquad i,j = 1,2,\ldots,k$$

is called an elementary block similarity transformation. We shall study the effect of F_{ij} on the ordered Schur form A. For $i > j$ we have using (2.1-3)

$$(2.1\text{–}7) \qquad \{A(E_i^T N_{ij} E_j)\}_{rs} = \sum_{p=s}^{r} A_{rp}(E_i^T N_{ij} E_j)_{ps}$$

$$= \begin{cases} 0, & s \neq j \\ A_{ri} N_{ij}, & s = j \end{cases}$$

$$(2.1\text{–}8) \qquad \{E_i^T N_{ij} E_j)A\}_{rs} = \sum_{p=s}^{r} (E_i^T N_{ij} E_j)_{rp} A_{ps}$$

$$= \begin{cases} 0, & r \neq i \\ N_{ij} A_{js}, & r = i \end{cases}$$

Triangularity of A yields

$$(2.1\text{–}9) \qquad (E_i^T N_{ij} E_j)\, A(E_i^T N_{ij} E_j) = E_i^T N_{ij}(E_j A E_i^T)N_{ij} E_j$$

$$= 0.$$

Using (2.1–9) with $i > j$ we have by a straightforward computation

$$[F_{ij}(A)]_{rs} = (E_{ij}^{-1} A E_{ij})_{rs}$$

$$= A_{rs} - \{(E_i^T N_{ij} E_j)A\}_{rs} + \{A(E_i^T N_{ij} E_j)\}_{rs}$$

Coupled with (2.1–7) and (2.1–8) this yields for $i > j$:

$$[F_{ij}(A)]_{rs} =$$

$$= \begin{cases} A_{is} - N_{ij} & r = i,\ s \neq j \quad (2.1\text{-}10) \\ A_{rs} + A_{ri} N_{ij} & r \neq i,\ s = j \quad (2.1\text{-}11) \\ A_{ij} + A_{ii} N_{ij} - N_{ij} A_{jj} & r = i,\ s = j \quad (2.1\text{-}12) \\ A_{rs} \qquad \text{otherwise} & (2.1\text{-}13) \end{cases}$$

Recalling that A is triangular we note that

$$[F_{ij}(A)]_{rs} = \begin{cases} 0 & r < s \\ A_{rr} & r = s \end{cases}$$

We note further that only elements in the ith row and those in the jth column of A are altered by F_{ij}.

For $i = j$ we have $E_{ii} = E_{ii}^{-1} = I$ and consequently $[F_{ii}(A)]_{rs} = A_{rs}$.

By (2.1-12) we see that $[F_{ij}(A)]_{ij} = 0$ if and only if there exists an $n_i \times n_j$ matrix N_{ij} such that

$$-A_{ii} N_{ij} + N_{ij} A_{jj} = A_{ij}. \quad (2.1\text{-}14)$$

That this equation is solvable may be seen from the following theorem which is proved in the Appendix.

THEOREM A-1. A necessary and sufficient condition that the matrix equation $-AX + XB = C$ have a solution for all C is that the eigenvalues of A be distinct from those of B.

Indeed, since A was assumed to be an ordered Schur form and $i \neq j$, the eigenvalues of A_{ii} are distinct from those of A_{jj}.

We shall now consider the effect of successive applications of elementary similarity transformations, F_{ij}, on an ordered Schur form A where at each stage N_{ij} is chosen as the solution to (2.1-14). For this we introduce the following notation.

$$\text{Let} \quad A^{(1,1)} = F_{11}(A) = A$$

$$A^{(2,1)} = F_{21}\left[A^{(1,1)}\right]$$

$$A^{(3,1)} = F_{31}\left[A^{(2,1)}\right]$$

$$\vdots$$

$$A^{(i,j)} = \begin{cases} F_{ij}\left[A^{(i-1,j)}\right], & i > j \\ F_{ii}\left[A^{(k,i-1)}\right] = A^{(k,i-1)}, & i = j = 2,3,\ldots,k \end{cases}$$

where F_{ij} is determined by the condition that

$$A_{ij}^{(i,j)} = 0, \quad \text{i.e., that (2.1-14) is satisfied.}$$

If we define $A^{(i-1,i)} = A^{(k,i-1)}$ for $i = 2,3,\ldots,k$ we may write

$$A^{(i,j)} = F_{ij}\left[A^{(i-1,j)}\right] \qquad i \geq j,\ i \neq 1$$

$$A^{(1,1)} = A.$$

Rewriting (2.1-10) thru (2.1-13) in our new notation we have

$$A_{rs}^{(i,j)}$$

$$= \begin{cases} A_{is}^{(i-1,j)} - N_{ij} A_{js}^{(i-1,j)} & r = i,\ s \neq j \quad (2.1\text{-}15) \\ A_{rj}^{(i-1,j)} + A_{ri}^{(i-1,j)} N_{ij} & r \neq i,\ s = j \quad (2.1\text{-}16) \\ A_{ij}^{(i-1,j)} + A_{ii}^{(i-1,j)} N_{ij} - N_{ij} A_{jj}^{(i-1,j)} & r = i,\ s = j \quad (2.1\text{-}17) \\ A_{rs}^{(i-1,j)} & \text{otherwise} \quad (2.1\text{-}18) \end{cases}$$

Since $A_{rr}^{(i,j)} = A_{rr}$ $(r = 1,2,\ldots,k)$ for all (i,j) we have

$$A_{ij}^{(i,j)} = A_{ij}^{(i-1,j)} + A_{ii} N_{ij} - N_{ij} A_{jj}. \quad (2.1\text{-}19)$$

We claim that the result of application of F_{ij} in some order is to reduce A to a quasi-diagonal form. That this is true can be seen from the following lemma.

LEMMA 7. If $A_{rs}^{(i-1,j)} = 0$ for $s > r$; for $s < j$ where $s < r \leq k$; and for $s = j$ where $j < r \leq i - 1$ and if N_{ij} is chosen such that $A_{ij}^{(i,j)} = 0$ in (2.1-19) then $A_{rs}^{(i,j)} = 0$ for $s > r$; for $s < j$ where $s < r \leq k$; and for $s = j$ where $j < r \leq i$.

Proof.

(i) $r \neq i,\ s \neq j$

$A_{rs}^{(i,j)} = A_{rs}^{(i-1,j)}$ by (2.1-18)

(ii) $r = i,\ s < j$.

By (2.1-15)

$$A_{is}^{(i,j)} = A_{is}^{(i-1,j)} - N_{ij} A_{js}^{(i-1,j)}$$

$$= 0$$

since $A_{is}^{(i-1,j)}$ and $A_{js}^{(i-1,j)}$ are, by hypothesis, null matrices.

(iii) $r = i, s > j.$

By (2.1-15)

$$A_{is}^{(i,j)} = A_{is}^{(i-1,j)} - N_{ij} A_{js}^{(i-1,j)}$$

$$= A_{is}^{(i-1,j)}$$

since $A^{(i-1,j)}$ is lower triangular.

(iv) $r < i, s = j.$

From (2.1-16) and the triangularity of $A^{(i-1,j)}$

$$A_{rj}^{(i,j)} = A_{rj}^{(i-1,j)} + A_{ri}^{(i-1,j)} N_{ij}$$

$$= A_{rj}^{(i-1,j)}$$

(v) $r > i, s = j.$

From (2.1-16)

$$A_{rj}^{(i,j)} = A_{rj}^{(i-1,j)} + A_{ri}^{(i-1,j)} N_{ij}$$

(vi) $r = i, s = j$

$$A_{ij}^{(i,j)} = 0$$

by definition of F_{ij}.

Comment: We have proved more than the statement of the lemma. Indeed we have shown that the only elements of $A^{(i-1,j)}$ which are altered by F_{ij} are $A_{kj}^{(i-1,j)}$, $k \geq i$.

Thus the sequence of elementary transformations $F_{21}, F_{31}, \ldots, F_{k1}$; $F_{32}, F_{42}, \ldots, F_{k2}; \ldots; F_{k,k-1}$ reduces A to a quasi-diagonal form whose diagonal blocks are precisely those of A.

Let X denote the matrix formed by the multiplication of the E_{ij} taken in the same order as the F_{ij}. Namely,

$$X = \prod_{j=1}^{k-1} \prod_{i>j} E_{ij}.$$

Then $X^{-1}AX = Q$, where Q is quasi-diagonal with $Q_{ii} = A_{ii}$. But from (2.1-5) and (2.1-6)

$$\prod_{j=1}^{k-1} \prod_{i>j} E_{ij} = N.$$

Thus $N = X$ and we have finally,

THEOREM 1. For a given ordered Schur matrix $A = (A_{ij})$ of order k, let $N = (N_{ij})$ be the sp-triangular matrix such that for $i > j$ N_{ij} satisfies

$$-A_{ii} N_{ij} + N_{ij} A_{jj} = A_{ij}^{(i-1,j)}$$

and such that

$$N_{ii} = I_{n_i} \qquad (i = 1,2,\ldots,k).$$

Then

$$N^{-1}AN = Q$$

where Q is quasi-diagonal with $Q_{ii} = A_{ii}$ $(i = 1,2,\ldots,k)$.

Section 2.2.

Introduction. We shall begin this section by determining an expression for the elements $A_{rj}^{(i-1,j)}$ in terms of elements of A and N. Using the integral representation formula for N_{ij} given in the Appendix, we are able to express N as a sum of certain matrices each of whose elements are integrals of certain functions of the elements of A.

Using the above representation of N we are able to give a bound for $C_2(N)$.

By further restricting the form of A we are able to give a bound for a condition number which does not require the explicit computation of N.

2.21. A representation for N.

We begin with the following lemmas.

LEMMA 8. For $j + p < r$,

$$A_{rj}^{(j+p,j)} = \sum_{\ell=0}^{p} A_{r,j+\ell} N_{j+\ell,j}, \quad N_{jj} = I_{nj}, \tag{2.2-1}$$

$$(j = 1,2,\ldots,k).$$

LEMMA 9. If $r > j$,

$$A_{rj}^{(r-1,j)} = \sum_{\ell=j}^{r-1} A_{r\ell} N_{\ell j}. \tag{2.2-2}$$

Lemma 9 is a particular case of Lemma 8 where $p = (r-1) - j$.

Proof of Lemma 8. For $p = 1$, we have by (2.1–16)

$$\begin{aligned} A_{rj}^{(j+1,j)} &= A_{rj}^{(j,j)} + A_{r,j+1}^{(j,j)} N_{j+1,j} \\ &= A_{rj}^{(k,j-1)} + A_{r,j+1}^{(k,j-1)} N_{j+1,j} \\ &= A_{rj} + A_{r,j+1} N_{j+1,j} \\ &= \sum_{\ell=0}^{1} A_{r,j+\ell} N_{j+\ell,j}. \end{aligned}$$

The validity of the next to last statement follows from Lemma 7. Assuming now that (2.2–1) holds for $p - 1$ we have by (2.1–16) and Lemma 7:

$$\begin{aligned} A_{rj}^{(j+p,j)} &= A_{rj}^{(j+p-1,j)} + A_{r,j+p}^{(j+p-1,j)} N_{j+p,j} \\ &= A_{rj}^{(j+p-1,j)} + A_{r,j+p} N_{j+p,j} \\ &= \sum_{\ell=0}^{p-1} A_{r,j+\ell} N_{j+\ell,j} + A_{r,j+p} N_{j+p,j} \\ &= \sum_{\ell=0}^{p} A_{r,j+\ell} N_{j+\ell,j} \end{aligned}$$

which is the statement of the lemma.

We now reproduce an integral representation for N_{ij}. The proof of the validity of this representation and related results are to be found in the Appendix.

THEOREM.

$$N_{ij} = -\int_0^\infty e^{-A_{ii}t} A_{ij}^{(i-1,j)} e^{A_{jj}t} dt.$$

Let now $\tilde{A} = A - dg(A)$. $\tilde{A}$ is then a strictly lower triangular matrix. Define now

$$\tilde{A}^{(0)} = I$$

$$\tilde{A}^{(1)} = (\tilde{A}^{(1)}_{rs}) \quad \text{where} \quad \tilde{A}^{(1)}_{rs} = -\int_0^\infty e^{-A_{rr}t} (\tilde{A}\,\tilde{A}^{(0)})_{rs} e^{A_{ss}t}\, dt,$$

$$\tilde{A}^{(2)} = (\tilde{A}^{(2)}_{rs}) \quad \text{where} \quad \tilde{A}^{(2)}_{rs} = -\int_0^\infty e^{-A_{rr}t} (\tilde{A}\,\tilde{A}^{(1)})_{rs} e^{A_{ss}t}\, dt,$$

or in general,

$$\tilde{A}^{(p+1)} = (\tilde{A}^{(p+1)}_{rs}); \quad \tilde{A}^{(p+1)}_{rs} = -\int_0^\infty e^{-A_{rr}t} (\tilde{A}\,\tilde{A}^{(p)})_{rs} e^{A_{ss}t}\, dt.$$

The above matrices will not be defined in all cases when the eigenvalues of A are complex. In this case we alter the definitions to read:

$$\tilde{A}^{(0)} = I$$

$$\tilde{A}^{(p+1)} = (\tilde{A}^{(p+1)}_{rs}); \quad \tilde{A}^{(p+1)}_{rs}$$

$$= -\int_0^\infty e^{i\theta} \exp\left[e^{-i\theta} A_{rr} t\right] (\tilde{A}\,\tilde{A}^{(p)})_{rs} \exp\left[e^{i\theta} A_{ss} t\right] dt$$

where θ is determined by the eigenvalues of A and is given by Lemma B of the Appendix. Noting that for any lower triangular matrix B, e^{Bt} is lower triangular we see that the diagonal entries of $\tilde{A}$ and hence of $\tilde{A}^{(1)}$ are all zero, the first two diagonals of $\tilde{A}^{(2)}$ vanish; indeed $\tilde{A}^{(r)} = 0$, $r \geq k$.

We are now able to represent N explicitly as indicated in the

introduction. Namely, we have the following theorem.

THEOREM 2. If

$$M = I + \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)} \tag{2.2-3}$$

then $N = M$.

Proof. $N_{jj} = I_{n_j}$ $(j = 1,2,\ldots,k)$ which agrees with the above. $M_{\ell j} = 0$ for $\ell < j$ according to (2.2-3). But $N_{\ell j} = 0$ for $\ell < j$ also, since N is lower triangular by construction. Assume now that

$$N_{\ell j} = (I + \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)})_{\ell j}; \qquad j < \ell \leq i.$$

Then, since

$$N_{i+1,j} = -\int_0^\infty e^{-A_{i+1,i+1}t} A_{i+1,j}^{(i,j)} e^{A_{jj}t}\, dt$$

and

$$A_{i+1,j}^{(i,j)} = \sum_{\ell=j}^{i} A_{i+1,\ell} N_{\ell j}$$

we have from Lemma 9

$$N_{i+1,j} = -\int_0^\infty e^{-A_{i+1,i+1}t} \left\{ \sum_{\ell=j}^{i} A_{i+1,\ell} N_{\ell j} \right\} e^{A_{jj}t}\, dt$$

$$= -\int_0^\infty e^{-A_{i+1,i+1}t} \left\{ \sum_{\ell=j}^{i} A_{i+1,\ell} (I_{\ell j} + \tilde{A}_{\ell j}^{(1)} + \cdots + \tilde{A}_{\ell j}^{(k-1)}) \right\} e^{A_{jj}t}\, dt$$

$$= -\int_0^\infty e^{-A_{i+1,i+1}t} \left\{ \sum_{\ell=j}^{i} \tilde{A}_{i+1,\ell} (I_{\ell j} + \tilde{A}_{\ell j}^{(1)} + \cdots + \tilde{A}_{\ell j}^{(k-1)}) \right\} e^{A_{jj}t}\, dt$$

$$= -\int_0^\infty e^{-A_{i+1,i+1}t} \left\{ (\tilde{A}\,\tilde{A}^{(0)})_{i+1,j} + (\tilde{A}\,\tilde{A}^{(1)})_{i+1,j} + \cdots \right.$$

$$\left. + (\tilde{A}\,\tilde{A}^{(k-2)})_{i+1,j} + (\tilde{A}\,\tilde{A}^{(k-1)})_{i+1,j} \right\} e^{A_{jj}t}\,dt$$

$$= \tilde{A}^{(1)}_{i+1,j} + \tilde{A}^{(2)}_{i+1,j} + \cdots + \tilde{A}^{(k-1)}_{i+1,j} + \tilde{A}^{(k)}_{i+1,j}$$

$$= (I + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)})_{i+1,j}$$

since $A_{rs} = \tilde{A}_{rs}$ for $r > s$; $\tilde{A}^{(k)} = 0$ and $I_{i+1,j} = 0$. The above induction step completes the proof.

It is appropriate at this point to consider the behavior of N as B approaches normality. Let us consider the given matrix B and $A = U^*BU$ where A is an ordered Schur form and U is unitary. We put

$$A = D + M,$$

where D denotes the diagonal matrix whose main diagonal coincides with that of A. Since ϵ is unitarily invariant, $[\epsilon(B)]^2 = [\epsilon(A)]^2 = [\epsilon(D)]^2 + [\epsilon(M)]^2$. It follows that

$$\epsilon(M) = \left\{ [\epsilon(B)]^2 - \sum_{i=1}^{n} |\lambda_i|^2 \right\}^{1/2}$$

is independent of the special choice of ordered Schur form. Noting that B is normal iff $\epsilon(M) = 0$ (see [19], Theorem 10.3.8), we see that B and, from the continuity of the Schur form, A approaches

normality as $\epsilon(M) \longrightarrow 0$ or, what is the same thing, as the off-diagonal elements of A approach zero. The elements of the matrices $\tilde{A}^{(i)}$ $i = 1,2,\ldots,k-1$ depend continuously on the off-diagonal elements of A and thus approach zero. Finally then, $N \longrightarrow I$ continuously as B aproaches normality.

2.22. A bound for the condition of N. It is now possible to give an upper bound for a condition number of N.

Let

$$L = \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)}.$$

Then L is strictly lower triangular and $L^r = 0$, $r \geq k$. Since $N = I + L$,

$$N^{-1} = (I + L)^{-1} = I - L + L^2 - \cdots (-1)^{k-1} L^{k-1}.$$

We have then the following theorem.

THEOREM 3. For any multiplicative norm ν,

$$C_\nu(N) \leq \nu(N) \left[\nu(I) + \nu(L) + \cdots + [\nu(L)]^{k-1}\right]$$

where $L = \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)}$.

2.23. Restricted Schur forms and a new bound for $C_\nu(N)$. Use of the bound for $C_\nu(N)$ given in Theorem 3 requires, of course, the calculation of $\tilde{A}^{(i)}$, $i = 1,2,\ldots,k-1$. Due to the prohibitive nature of the calculations required we shall derive a new bound for $C_\nu(N)$. This bound does not require the calculation of the $\tilde{A}^{(i)}$ but further restricts the Schur form and does not in general yield

as sharp a bound as that given by Theorem 3.

We begin by finding a bound for the norm of X, where

$$X = -\int_0^\infty e^{-At} C e^{Bt}$$

and A and B are lower triangular. It will be necessary to make certain assumptions regarding the eigenvalues of A and B. We need however, to prove the following lemma first.

LEMMA 10. *Let $B(t)$ be any Riemann integrable matrix function (i.e., each element of $B(t)$ is integrable), and let $||\cdot||$ denote an arbitrary matrix norm.*

If

$$A = \int_0^\infty B(t)dt$$

then

$$||A|| \leq \int_0^\infty ||B(t)||dt.$$

Proof. Let

$$A(x) = \int_0^x B(t)dt.$$

Then

$$A'(x) = B(x) \quad \text{and} \quad A(\infty) = A.$$

By a result of Dahlquist [7] we have for every matrix $A(x)$ and norm $||\cdot||$ that

$$||A(x)||' \leq ||A'(x)||. \tag{2.2-4}$$

Integrating (2.2-4)

$$||A(x)|| - ||A(0)|| \leq \int_0^x ||A'(t)||dt$$

or

$$||A(x)|| \leq \int_0^x ||B(t)||dt.$$

Hence

$$||A|| \leq \int_0^\infty ||B(t)||dt.$$

We now turn to the problem of finding a bound for the norm of X;

$$X = -\int_0^\infty e^{-At} C e^{Bt} dt.$$

We shall assume that

(1) A, B are lower triangular matrices of order n_A, n_B respectively.

(2) The diagonal entries of A(B) are all equal to some constant we shall denote by λ_A (λ_B). i.e. A(B) has only one eigenvalue λ_A (λ_B) repeated n_A (n_B) times.

(3) $\lambda_A > \lambda_B$

A and B can then be written as

$$A = \lambda_A I + L_A$$
$$B = \lambda_B I + L_B$$

where L_A and L_B are strictly lower triangular.

Since $\lambda_A I$ and $\lambda_B I$ are scalar matrices they commute respectively with L_A and L_B, and we have

$$e^{-At} = e^{-(\lambda_A I + L_A)t} = e^{-\lambda_A I t}\, e^{-L_A t}$$

$$e^{Bt} = e^{(\lambda_B I + L_B)t} = e^{\lambda_B I t}\, e^{L_B t}$$

$$e^{-At}\, Ce^{Bt} = e^{(\lambda_B - \lambda_A)It} \sum_{r=0}^{n_A-1} \sum_{s=0}^{n_B-1} (-1)^r \frac{L_A^r}{r!}\, C\, \frac{L_B^s}{s!}\, t^{r+s}.$$

For any axis-oriented multiplicative norm $|||\cdot|||$ we have from Lemma 10 that

$$|||X||| \le \int_0^\infty e^{(\lambda_B - \lambda_A)t} \sum_{r=0}^{n_A-1} \sum_{s=0}^{n_B-1} \frac{1}{r!\,s!}\, |||L_A|||^r\, |||C|||\, |||L_B|||^s\, t^{r+s}\, dt$$

$$= |||C||| \sum_{r=0}^{n_A-1} \sum_{s=0}^{n_B-1} \left\{ \frac{1}{r!\,s!}\, |||L_A|||^r\, |||L_B|||^s \int_0^\infty e^{(\lambda_B - \lambda_A)t}\, t^{r+s}\, dt \right\}$$

$$= \frac{|||C|||}{\lambda_A - \lambda_B} \sum_{r=0}^{n_A-1} \sum_{s=0}^{n_B-1} \left\{ \frac{1}{r!\,s!} \left(\frac{|||L_A|||}{\lambda_A - \lambda_B} \right)^r \left(\frac{|||L_B|||}{\lambda_A - \lambda_B} \right)^s \int_0^\infty e^{-t}\, t^{r+s}\, dt \right\}$$

$$= \frac{|||C|||}{\lambda_A - \lambda_B} \sum_{r=0}^{n_A-1} \sum_{s=0}^{n_B-1} \binom{r+s}{r} \left(\frac{|||L_A|||}{\lambda_A - \lambda_B} \right)^r \left(\frac{|||L_B|||}{\lambda_A - \lambda_B} \right)^s$$

Note that

$$\sum_{r=0}^{R} \sum_{s=0}^{S} \binom{r+s}{s} x^r y^s \le \sum_{k=0}^{R+S} k! \sum_{p=0}^{k} \frac{x^p y^{k-p}}{p!\,(k-p)!}$$

$$= \sum_{k=0}^{R+S} \sum_{p=0}^{k} \binom{k}{p} x^p y^{k-p}$$

$$= \sum_{k=0}^{R+S} (x+y)^k.$$

We have, upon substituting $\ell_A = \|L_A\|$, $\ell_B = \|L_B\|$

$$\|X\| \leq \frac{\|C\|}{\lambda_A - \lambda_B} \sum_{k=0}^{n_A+n_B-2} \left(\frac{\ell_A + \ell_B}{\lambda_A - \lambda_B}\right)^k$$

or

$$\|X\| \leq \frac{\|C\|}{\lambda_A - \lambda_B} \sum_{k=0}^{n_A+n_B-2} q_{AB}^k \tag{2.2-5}$$

$$q_{AB} = \frac{\ell_A + \ell_B}{\lambda_A - \lambda_B} .$$

This matrix X is, of course, the solution of the matrix equation $-AX + XB = C$ provided $\lambda_A > \lambda_B$. If however we only know that $\lambda_A \succ \lambda_B$ the solution to the matrix equation is given (see Lemma B, Appendix) by

$$X = -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} At\right] C \exp\left[e^{i\theta} Bt\right] dt$$

where θ is given by the above cited theorem. We may then generalize (2.2-5) in the case where $\lambda_A \succ \lambda_B$.

LEMMA 11. Let A, B be lower triangular matrices of order n_A, n_B respectively, with repeated roots λ_A, λ_B respectively. If $\lambda_A \succ \lambda_B$ and

$$X = -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} At\right] C \exp\left[e^{i\theta} Bt\right] dt$$

is a solution of the matrix equation $-AX + XB = C$ then

$$\|X\| \leq \frac{\|C\|}{|\lambda_A - \lambda_B|} \sum_{k=0}^{n_A+n_B-2} q_{AB}^k$$

$$q_{AB} = \frac{\ell_A + \ell_B}{|\lambda_A - \lambda_B|} .$$

The proof is similar to that used in deriving (2.2-5) in which $|\lambda_A - \lambda_B|$ replaces $\lambda_A - \lambda_B$ and will not be repeated.

DEFINITION. A restricted Schur form is an ordered Schur form in which the sets S_i introduced in section 2.1 each contain only (repeated) eigenvalue.

Let $A = (A_{ij})$ be such a form. Then we may write

$$A_{rr} = \lambda_r I + L_r$$

$$A_{ss} = \lambda_s I + L_s$$

where A_{rr} (A_{ss}) is of order $n_r(n_s)$ and L_r and L_s are strictly lower triangular matrices. If $\lambda_r \succ \lambda_s$ let

$$\gamma_{rs} = \frac{1}{|\lambda_r - \lambda_s|} \sum_{k=0}^{n_r+n_s-2} q_{rs}^k$$

where

$$q_{rs} = \frac{\ell_r + \ell_s}{|\lambda_r - \lambda_s|}$$

and $\ell_r = \|L_r\|$, $\ell_s = \|L_s\|$ for an axis oriented multiplicative norm, $\|\cdot\|$. From Lemma 11,

$$\left\| -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} A_{rr} t\right] C\left[e^{i\theta} A_{ss} t\right] dt \right\| \leq \gamma_{rs} \|C\|.$$

Since

$$\tilde{A}_{rs}^{(p)} = -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} A_{rr} t\right] (\tilde{A}\,\tilde{A}^{(p-1)})_{rs} \exp\left[e^{i\theta} A_{ss} t\right] dt$$

$$\|\tilde{A}_{rs}^{(p)}\| \leq \gamma_{rs} \|(\tilde{A}\,\tilde{A}^{(p-1)})_{rs}\|.$$

For $r > s$, let

$$(2.2\text{-}6)\quad Y_{rs}^{(p)} = \sum_{s<n_1<n_2<\cdots<n_{p-1}<r} \|A_{rn_{p-1}}\| \, \|A_{n_{p-1}n_{p-2}}\| \cdots \|A_{n_1 s}\| \, \gamma_{rs}\gamma_{n_{p-1}s} \cdots \gamma_{n_1 s} \qquad p = 1,2,\ldots,k-1$$

$$Y_{rs}^{(0)} = \|I_{rs}\|.$$

For $r \leq s$, let

$$Y_{rs}^{(p)} = 0 \qquad (p = 1,2,\ldots,k-1)$$

$$Y_{rs}^{(0)} = \|I_{rs}\|.$$

LEMMA 12. For $r > s$, $p = 0,1,\ldots,k-1$

$$||\tilde{A}^{(p)}_{rs}|| \leq Y^{(p)}_{rs}.$$

Proof.

$$||\tilde{A}^{(0)}_{rs}|| = ||I_{rs}||$$

$$||A^{(1)}_{rs}|| = \left|\left|-\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} A_{rr} t\right] (\tilde{A}\,\tilde{A}^{(0)})_{rs} \exp\left[e^{i\theta} A_{ss} t\right] dt\right|\right|$$

$$\leq \gamma_{rs} ||\tilde{A}_{rs}||$$

$$= \gamma_{rs} ||A_{rs}||$$

$$= Y^{(1)}_{rs}.$$

Assuming the inequality holds for p,

$$||\tilde{A}^{(p+1)}_{rs}|| = \left|\left|-\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} A_{rr} t\right] (\tilde{A}\,\tilde{A}^{(p)})_{rs} \exp\left[e^{i\theta} A_{ss} t\right] dt\right|\right|$$

$$\leq \gamma_{rs} ||(\tilde{A}\,\tilde{A}^{(p)})_{rs}||$$

$$\leq \gamma_{rs} \left|\left| \sum_{s<n_p<r} A_{rn_p} \tilde{A}^{(p)}_{n_p s} \right|\right|$$

$$\leq \gamma_{rs} \sum_{s<n_p<r} ||A_{rn_p}|| \, ||\tilde{A}^{(p)}_{n_p s}||$$

$$\leq \gamma_{rs} \sum_{s<n_p<r} ||A_{rn_p}|| \sum_{s<n_1<n_2<\cdots<n_p} ||A_{n_p, n_{p-1}}||$$

$$||A_{n_{p-1}, n_{p-2}}|| \cdots ||A_{n_1 s}|| \, \gamma_{n_p s} \, \gamma_{n_{p-1} s} \cdots \gamma_{n_1 s}$$

$$= \gamma_{rs} \sum_{s<n_1<n_2<\cdots<n_p<r} ||A_{rn_p}|| \, ||A_{n_{p-1}, n_{p-2}}|| \cdots$$

$$\leq ||A_{n_1 s}|| \gamma_{n_p s} \gamma_{n_{p-1} s} \cdots \gamma_{n_1 s}$$

$$= Y_{rs}^{(p+1)}$$

which is the statement of the lemma for $p + 1$.

Define

$$\hat{A}^{(p)} = (||\tilde{A}_{rs}^{(p)}||)$$

$$Y^{(p)} = (Y_{rs}^{(p)})$$

$$\hat{N} = (||N_{rs}||)$$

for $r,s = 1,2,\ldots,k, \quad p = 0,1,\ldots,k-1.$

Then, recalling that

$$N = I + \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)},$$

we have

$$\hat{N} \leq I_k + \hat{A}^{(1)} + \hat{A}^{(2)} + \cdots + \hat{A}^{(k-1)}$$

$$\leq I_k + Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)}$$

for $\hat{A}^{(p)} \leq Y^{(p)}$ in view of Lemma 12.

If ν is a monotone norm

$$\nu(\hat{N}) \leq \nu(I_k + Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)}).$$

But, as shown in Lemma 3, the function f defined by

$f(M) = \nu(\hat{M})$ for all $k \times k$ partitioned matrices M, is a norm.

We have then the following result.

THEOREM 4. *Let $A = (A_{ij})$ be a restricted Schur form, $\|\cdot\|$ a multiplicative axis-oriented matrix norm and ν a monotone norm. Then the function f defined above is a norm and*

$$f(N) \leq \nu(I_k + Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)})$$

where the elements of $Y^{(p)}$ are given by (2.2-6).

Writing

$$N = I + P$$

with

$$P = \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)},$$

$$N^{-1} = (I + P)^{-1} = I - P + P^2 - \cdots (-1)^{k-1} P^{k-1}$$

$$\widehat{[N^{-1}]} \leq I_k + \hat{P} + \widehat{[P^2]} + \cdots + \widehat{[P^{k-1}]} \tag{2.2-7}$$

$$\leq I_k + \hat{P} + (\hat{P})^2 + \cdots + (\hat{P})^{k-1}$$

since $\widehat{|P^r|} \leq (\hat{P})^r$ for all r.

Noting that $\hat{P} \leq \hat{A}^{(1)} + \hat{A}^{(2)} + \cdots + \hat{A}^{(k-1)}$

$$\leq Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)}$$

and setting $L = Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)}$ we have from (2.2-7)

$$\widehat{[N^{-1}]} \leq I_k + L + L^2 + \cdots + L^{(k-1)}$$

For any monotone norm ν,

$$\nu\widehat{[N^{-1}]} \leq \nu(I_k) + \nu(L) + [\nu(L)]^2 + \cdots + [\nu(L)]^{k-1}.$$

Setting $f(N^{-1}) = \nu\widehat{[N^{-1}]}$ we have by Lemma 3 that f is a norm of N^{-1} and consequently arrive at the following theorem.

THEOREM 5. *With the same hypothesese as in Theorem 4 we have*

$$C_f(N) = f(N)\, f(N^{-1})$$

$$\leq \nu(I_k + L)\left[\nu(I_k) + \nu(L) + [\nu(L)]^2 + \cdots + [\nu(L)]^{k-1}\right]$$

where $L = Y^{(1)} + Y^{(2)} + \cdots + Y^{(k-1)}$, *the* $Y^{(p)}$ *defined as in* (2.2–6).

CHAPTER 3

Introduction. The representation for N given here is identical to that given in the preceeding chapter. This follows from the uniqueness of the solution of the matrix equation $-AX + XB = C$ guaranteed by the assumption of an ordered Schur form. The preceeding chapter has demonstrated the formal methods for block elimination and at the same time indicated the complicated process of determining N. Computationally it will be seen that the methods of this chapter are superior to those of Chapter 2. Proofs given in that chapter shall be adopted here specializing to the case in hand. The development of this chapter is more straightforward than that of the preceeding and is recommended for practical applications. Chapter 2 should then primarily be considered for its theoretical value.

Section 3.1. A scalar development for N.

We present here a scalar analogue to the material developed in 2.1. We again assume that $A = (A_{ij})$, $i,j = 1,2,\ldots,k$ is an ordered Schur form with A_{ij} of order $n_i \times n_j$ with $\sum_{i=1}^{k} n_i = n$. Letting a_{ij} denote the (i,j) element of A considered as a scalar matrix, we say that the pair of indices (i,j) with $i > j$ is of type P and denote this by $(i,j) \in P$ if a_{ij} is not contained in any of the diagonal blocks of A considered as an ordered Schur form.

If e_i denotes the $1 \times n$ row vector whose only non-zero element appears in the ith position it follows that

$$e_i e_j^T = e_j e_i^T = \begin{cases} 0 & j \neq i \\ 1 & j = i \end{cases}$$

and

$$a_{ij} = e_i A e_j^T$$

DEFINITION. The matrices $\mathcal{E}_{ij}$ defined by

$$\mathcal{E}_{ij} = \mathcal{E}_{ij}[e_i, e_j, n_{ij}] = I + e_i^T n_{ij} e_j$$

where n_{ij} is an arbitrary complex number for $(i,j) \in P$ and zero otherwise are called P-elementary matrices.

$$(\mathcal{E}_{ij})_{rs} = \begin{cases} 1 & r = s \\ n_{ij} & r = i, s = j \\ 0 & \text{otherwise} \end{cases}$$

The following are immediate results of the above definitions.

$$\mathcal{E}_{ij}^{-1}[e_i, e_j, n_{ij}] = \mathcal{E}_{ij}[e_i, e_j, -n_{ij}].$$

If

$$N_j = \prod_{i=1}^{n} \mathcal{E}_{ij} = \prod_{i>j} \mathcal{E}_{ij} \tag{3.1-1}$$

$$N_j = I + \Big(\sum_{i>j} e_i^T n_{ij} \Big) e_j$$

That is

$$(N_j)_{rs} = \begin{cases} 1 & r = s = i \\ n_{rj} & s = j \\ 0 & \text{otherwise} \end{cases}$$

Furthermore

$$N_j^{-1} = I - \Big(\sum_{i>j} e_i^T n_{ij} \Big) e_j .$$

Letting

$$N = \prod_{j=1}^{n} N_j \tag{3.1-2}$$

we have

$$N_{rs} = \begin{cases} 1 & r = s = i \\ n_{rs} & \text{otherwise} \end{cases}$$

We proceed, as in Chapter 2, to eliminate one by one all those elements whose subscripts are of type P. In this manner we shall again arrive at a quasi-diagonal matrix.

DEFINITION. The function f_{ij} defined on all matrices A of order n by

$$f_{ij}(A) = \mathcal{E}_{ij}^{-1} A \mathcal{E}_{ij}$$

is called a P elementary similarity transformation.

Let

$$A' = f_{ij}(A).$$

Using results similar to (2.1–7), (2.1–8 and (2.1–9) we have

$$a'_{rs} = \begin{cases} a_{is} - n_{ij}\, a_{js} & r = i,\ s \neq j \quad (3.1\text{–}3) \\ a_{rj} + a_{ri}\, n_{ij} & r \neq i,\ s = j \quad (3.1\text{–}4) \\ a_{ij} + a_{ii}\, n_{ij} - n_{ij}\, a_{jj} & r = i,\ s = j \quad (3.1\text{–}5) \\ a_{rs} & \text{otherwise} \quad (3.1\text{–}6) \end{cases}$$

Only elements in the ith row and those in the jth column whose subscripts are of type P are affected by f_{ij} and $A'_{rr} = A_{rr}$, $r = 1, 2, \ldots, k$.

By (3.1–5), $a'_{ij} = 0$ if and only if there exists a complex number n_{ij} such that

$$-a_{ii}\, n_{ij} + n_{ij}\, a_{jj} = a_{ij}. \quad (3.1\text{–}7)$$

That this is always possible can be seen by choosing

$$n_{ij} = \frac{a_{ij}}{a_{jj} - a_{ii}}.$$

The denominator never vanishes, for $(i,j) \in P$ which precludes the possibility that $a_{ii} = a_{jj}$.

We now study the effect of successive applications of P-elementary similarity transformations, f_{ij}, where at each stage $f_{ij}\,[n_{ij}]$ is chosen such that equations similar to (3.1-7) are satisfied.

Let

$$A^{(1,1)} = f_{11}(A) = A$$

$$\vdots$$

$$A^{(i,j)} = f_{ij}\left[A^{(i-1,j)}\right]$$

where f_{ij} is chosen such that $a_{ij}^{(i,j)} = 0$ in (3.1-7). We may then write

$$A^{(i,j)} = \mathcal{E}_{ij}^{-1}\, A^{(i-1,j)}\, \mathcal{E}_{ij}, \qquad (i,j) \neq (1,1)$$

$$A^{(1,1)} = A$$

Rewriting (3.1-3) through (3.1-6) in our new notation we have

$$(3.1\text{-}8) \qquad a_{rs}^{(i,j)} = \begin{cases} a_{is}^{(i-1,j)} - n_{ij}\, a_{rs}^{(i-1,j)} & r = i,\ s \neq j \\ a_{rs}^{(i-1,j)} + a_{ri}^{(i-1,j)}\, n_{ij} & r \neq i,\ s = j \\ a_{ij}^{(i-1,j)} + a_{ii}^{(i-1,j)}\, n_{ij}\, a_{jj}^{(i-1,j)} & r = i,\ s = j \\ a_{rs}^{(i-1,j)} & \text{otherwise} \end{cases}$$

Since $a_{rr}^{(i,j)} = a_{rr}$, $r = 1,2,\ldots,n$, we have

$$a_{ij}^{(i,j)} = a_{ij}^{(i-1,j)} + a_{ii}\, n_{ij} - n_{ij}\, a_{jj} \tag{3.1-9}$$

We claim that the result of application of f_{ij} in some order is to reduce A to a quasi-diagonal form. This follows from the next lemma.

LEMMA 13. *Suppose we have* $a_{rs}^{(i-1,j)} = 0$ *for* $s > r$; $s < j$ *with* $s < r \leq n$; $s = j$ *with* $j < r \leq i-1$ *and let* n_{ij} *be chosen such that* $a_{ij}^{(i,j)} = 0$ *in* (3.1-9); *then* $a_{rs}^{(i-1,j)} = 0$ *for* $s > r$; $s < j$ *with* $s < r \leq n$; $s = j$ *with* $j < r \leq i$.

The proof is analogous to that of Lemma 7 of Chapter 2 and will not be repeated. As in Lemma 7 the proof shows that the only elements of $A^{(i-1,j)}$ which are altered by f_{ij} are $[A^{(i-1,j)}]_{kj}$, $k \geq i$ with $(k,j) \in P$.

Thus the sequence of elementary transformations, determined by eliminating each element with indices of type P, progressing down each column first and then by columns left to right, reduces A to a quasi-diagonal matrix whose diagonal entries are precisely those of A.

Let X be the matrix formed by the multiplication of the $\mathcal{E}_{ij}$ taken in the same order as the f_{ij}. Namely

$$X = \prod_{j=1}^{n} \prod_{i>j} \mathcal{E}_{ij}.$$

Then $X^{-1}AX = Q$, a quasi-diagonal matrix with $Q_{ii} = A_{ii}$. But from (3.1-1) and (3.1-2)

$$\prod_{j=1}^{n} \prod_{i>j} \mathcal{E}_{ij} = N.$$

Thus $X = N$ and we have:

THEOREM 6. *Let* $N = (n_{ij})$ *be a triangular matrix such that* n_{ij} *satisfies* $-a_{ii}\, n_{ij} + n_{ij}\, a_{jj} = a_{ij}^{(i-1,j)}$ *for* $(i,j) \in P$; $n_{ij} = 0$ *for* $(i,j) \notin P$ *except that* $n_{ii} = 1$, $i = 1,2,\ldots,n$; *then*

$$N^{-1}AN = Q$$

where Q *is quasi-diagonal and* $Q_{ii} = A_{ii}$ $i = 1,2,\ldots,k$.

Section 3.2. A bound for a condition number of N.

This section is devoted to the determination of a bound for a condition number of N, where N was defined in the preceeding section. We begin by finding an expression for the element of the form $a_{rj}^{(r-1,j)}$ which involves elements of A and N. Finally we find an expression for N which involves only elements of A.

LEMMA 14. For $j + p < r$,

$$a_{rj}^{(j+p,j)} = \sum_{\ell=0}^{p} a_{r,j+\ell}\, n_{j+\ell,j}. \tag{3.2-1}$$

Proof. We shall prove this lemma by induction on p.

For $p = 1$ we have by (3.1-8)

$$\begin{aligned} a_{rj}^{(j+1,j)} &= a_{rj}^{(j,j)} + a_{r,j+1}^{(j,j)}\, n_{j+1,j} \\ &= a_{rj}^{(n,j-1)} + a_{r,j+1}^{(n,j-1)}\, n_{j+1,j} \\ &= a_{rj} + a_{r,j+1}\, n_{j+1,j} \\ &= \sum_{\ell=0}^{1} a_{r,j+\ell}\, n_{j+\ell,j} \end{aligned} \tag{3.2-2}$$

(3.2-2) holding by virtue of Lemma 13.

Assuming (3.2-1) holds for $p - 1$ we have by (3.1-8)

$$a_{rj}^{(j+p,j)} = a_{rj}^{(j+p-1,j)} + a_{r,j+p}^{(j+p-1,j)}\, n_{j+p,j}$$

$$= a_{rj}^{(j+p-1,j)} + a_{r,j+p}\, n_{j+p,j} \qquad (3.2\text{–}3)$$

$$= \sum_{\ell=0}^{p-1} a_{r,j+\ell}\, n_{j+\ell,j} + a_{r,j+p}\, n_{j+p,j}$$

$$= \sum_{\ell=0}^{p} a_{r,j+\ell}\, n_{j+\ell,j}$$

(3.2–3) holding by virtue of Lemma 13.

In particular, we have for $p = (r - i) - j$

LEMMA 15. *If* $r > j$,

$$a_{rj}^{(r-1,j)} = \sum_{\ell=j}^{r-1} a_{r\ell}\, n_{\ell j}.$$

As we have seen before

$$n_{ij} = \frac{a_{ij}^{(i-1,j)}}{a_{jj} - a_{ii}} \quad \text{for} \quad (i,j) \in P.$$

Recalling that the diagonal entries a_{jj} of A are its eigenvalues we may write λ_j for a_{jj} and λ_i for a_{ii}. Then

$$n_{ij} = \frac{a_{ij}^{(i-1,j)}}{\lambda_j - \lambda_i}.$$

Let

$$\sigma_{ij} = \frac{1}{\lambda_j - \lambda_i}$$

and

$$\tilde{A} = A - \text{diag}\,(A_1, A_2, \ldots, A_k)$$

$\tilde{A}$ is then a (strictly) lower triangular scalar matrix such that

$$\tilde{a}_{ij} = \begin{cases} a_{ij} & \text{for } (i,j) \in P \\ 0 & \text{otherwise} \end{cases}$$

Define

$$\tilde{A}^{(0)} = I$$

$$\tilde{A}^{(1)} : \tilde{A}^{(1)}_{ij} = (\tilde{A}\,\tilde{A}^{(0)})_{ij}\,\sigma_{ij}$$

and in general

$$\tilde{A}^{(r)} : \tilde{A}^{(r)}_{ij} = (\tilde{A}\,\tilde{A}^{(r-1)})_{ij}\,\sigma_{ij}.$$

Note that $\tilde{A}^{(r)} = 0$, $r \geq k$ since each successive multiplication by $\tilde{A}$ introduces at least one new diagonal of zero block matrices.

THEOREM 7.

$$N = \tilde{A}^{(0)} + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)}.$$

Proof. $n_{jj} = 1$ by construction and $(\tilde{A}^{(0)} + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)})_{ij} = 0$ for $(i,j) \notin P$, which agrees with the above.

Assume now that

$$n_{\ell j} = (\tilde{A}^{(0)} + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)})_{\ell j} \quad \text{for} \quad j < \ell \leq i, \quad (\ell, j) \in P$$

then

$$n_{i+1,j} = \frac{a^{(i,j)}_{i+1,j}}{\lambda_j - \lambda_{i+1}} = a^{(i,j)}_{i+1,j}\ \sigma_{i+1,j}$$

$$= \left(\sum_{\ell=j}^{i} a_{i+1,\ell}\ n_{\ell j} \right) \sigma_{i+1,j}$$

$$= \left(\sum_{\ell=j}^{i} \tilde{a}_{i+1,\ell}\ n_{\ell j} \right) \sigma_{i+1,j}$$

$$= \left\{ \sum_{\ell=j}^{i} \tilde{a}_{i+1,\ell}\ (\tilde{A}^{(0)} + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)})_{\ell j} \right\} \sigma_{i+1,j}$$

$$= \sum_{\ell=j}^{i} \tilde{a}_{i+1,\ell}\ (\tilde{A}^{(0)}_{\ell j} + \tilde{A}^{(1)}_{\ell j} + \cdots + \tilde{A}^{(k-1)}_{\ell j})\ \sigma_{i+1,j}$$

$$= \left\{ (\tilde{A}\ \tilde{A}^{(0)})_{i+1,j} + (\tilde{A}\ \tilde{A}^{(1)})_{i+1,j} + \cdots + (\tilde{A}\ \tilde{A}^{(k-1)})_{i+1,j} \right\} \sigma_{i+1,j}$$

$$= \tilde{A}^{(1)}_{i+1,j} + \tilde{A}^{(2)}_{i+1,j} + \cdots + \tilde{A}^{(k-1)}_{i+1,j} + \tilde{A}^{(k)}_{i+1,j}$$

$$= (\tilde{A}^{(0)} + \tilde{A}^{(1)} + \cdots + \tilde{A}^{(k-1)})_{i+1,j}$$

since $\tilde{A}^{(k)} = 0$ and $\tilde{A}^{(0)}_{i+1,j} = 0$.

Letting $L = \tilde{A}^{(1)} + \tilde{A}^{(2)} + \cdots + \tilde{A}^{(k-1)}$ we note that L is strictly lower triangular and $L^r = 0$, $r \geq k$. Thus

$$N^{-1} = (I + L)^{-1} = I - L + L^2 - \cdots (-1)^{k-1} L^{k-1}.$$

THEOREM 8. If ν is any multiplicative norm,

$$C_\nu(N) \leq \nu(I + L)\{\nu(I) + \nu(L) + \cdots + [\nu(L)]^{k-1}\}.$$

Inasmuch as the elements of $\tilde{A}^{(r)}$ and hence L are easily computable it is not necessary to introduce a restricted Schur form.

CHAPTER 4

INTRODUCTION. Two applications of the results of the preceeding chapters will be considered here. Section 4.1 will deal with bounds for norms of powers of a fixed non-normal matrix. Bounds for the norm of the inverse of a fixed matrix are developed in Section 4.2 and applied to the problem of estimating residual vectors and matrices associated with the approximate solution of linear systems and approximate inverses.

Section 4.1. Iterated Matrices. The interest for bounds of norms of certain matrices arises principally from the study of finite difference schemes for solving hyperbolic and parabolic differential equations. Such bounds have been given by Lax and Richtmeyer [18]. For arbitrary matrices Gautschi [9, 10] and Ostrowski [24] have developed estimates which require some knowledge of the Jordan canonical form. More recently Henrici [14] has given bounds which depend upon the spectral radius and a certain measure of non-normality introduced in his paper.

For normal matrices we have of course

$$\sigma(B^r) = \lambda_B^r \tag{4.0-1}$$

as a simple consequence of the fact that normal matrices are unitarily similar to diagonal matrices and σ is both a unitarily invariant and an axis-oriented norm.

In contrast to the above results for non-normal matrices, Theorem 9 below gives an estimate for $\sigma(B^r)$ which depends upon all the eigenvalues of B according to their multiplicities, and a condition number. These results reduce to (4.0-1) for B normal.

Let then B be a given $n \times n$ matrix and U, a unitary matrix such that

$$A = UBU^*$$

is an ordered Schur form. Let N be chosen as in Theorem 6 such that $Q = N^{-1}AN$ is quasi-diagonal with $Q = \text{diag}\,(Q_{11}, Q_{22}, \ldots, Q_{kk})$ and $Q_{ii} = A_{ii}$ of order n_i, $i = 1, 2, \ldots, k$.

Thus

$$A^r = NQ^rN^{-1}$$

$$= N \,\text{diag}\,(Q_{11}^r, Q_{22}^r, \ldots, Q_{kk}^r)N^{-1}.$$

If σ represents the spectral norm, we have using Lemma 6

$$\sigma(A^r) \leq C_\sigma(N)\ \sigma\{\text{diag}\,(Q_{11}^r, Q_{22}^r, \ldots, Q_{kk}^r)\} \tag{4.1-1}$$

$$\leq C_\sigma(N) \max_{1 \leq i \leq k} \sigma(Q_{ii}^r)$$

Let Q_{ii} be written as a sum of a diagonal matrix D_i containing only its diagonal terms, and a strictly lower triangular matrix, L_i.

$$Q_{ii} = D_i + L_i$$

and

$$L_i^r = 0, \quad r \geq n_i .$$

For an arbitrary ordered Schur form the D_i are not necessarily scalar matrices and consequently will not commute with the L_i, preventing us from expanding $(D_i + L_i)^r$ according to the binomial theorem. But since D_i is diagonal, if we expand Q_{ii}^r any term with more than $n_i - 1$ L_i's vanishes. Thus

$$(4.1\text{–}2) \quad \sigma(Q_{ii}^r) \leq \Delta_i^r + \binom{r}{1} \Delta_i^{r-1} \ell_i + \cdots + \binom{r}{n_i - 1} \Delta_i^{r-n_i+1} \ell_i^{n_i-1}$$

where $\ell_i = \sigma(L_i)$ and $\Delta_i = \lambda_{Q_{ii}}$

From (4.1–1) and (4.2–2),

$$\sigma(A^r) \leq C_\sigma(N) \max_{1 \leq i \leq k} \left\{ \Delta_i^r + \binom{r}{1} \Delta_i^{r-1} \ell_i + \cdots + \binom{r}{n_i - 1} \Delta_i^{r-n_i+1} \ell_i^{n_i-1} \right\} .$$

Noting that $B^r = U^* A^r U$ and recalling that σ is unitarily invariant we have:

$$\sigma(B^r) \leq C_\sigma(N) \max_{1 \leq i \leq k} \left\{ \Delta_i^r + \binom{r}{1} \Delta_i^{r-1} \ell_i + \cdots + \binom{r}{n_i - 1} \Delta_i^{r-n_i+1} \ell_i^{n_i-1} \right\} .$$

The estimate holds for any ordered Schur form. Indeed, since specification of the ordering of the eigenvalues does not uniquely determine a Schur form we can conclude:

THEOREM 9. If $\lambda_B > 0$,

$$\sigma(B^r) \le \min\left[C_\sigma(N) \max_i \left\{ \Delta_i^r + \binom{r}{1} \Delta_i^{r-1} \ell_i + \cdots + \binom{r}{n_i - 1} \Delta_i^{r-n_i+1} \ell_i^{n_i-1} \right\}\right].$$

If $\lambda_B = 0$,

$$\sigma(B^r) \le \min\left[C_\sigma(N) \max_i \ell_i^r\right], \qquad r = 0,1,\ldots,M-1$$

$$\sigma(B^r) = 0, \qquad r \ge M$$

where $M = \max n_i$ and where the minimum is taken over all ordered Schur forms.

If B were normal, any Schur form would be diagonal implying that $N = I$. Thus $C_\sigma(N) = 1$ and $\sigma(B^r) \le \max \Delta_i^r = \lambda_B^r$. Since $\lambda_B^r \le \sigma(B^r)$, $\sigma(B^r) = \lambda_B^r$ in agreement with (4.0–1).

Section 4.2. Bounds for inverses. Let B be an arbitrary non-singular matrix, b a given vector and $\tilde{x}$ an alleged solution of $Bx = b$. If we define the residual of $\tilde{x}$ by $r = B\tilde{x} - b$, and if ν is a vector norm, the error $\tilde{x} - B^{-1}b = B^{-1}r$ of $\tilde{x}$ can be estimated as follows:

$$\nu(\tilde{x} - B^{-1}b) = \nu(B^{-1}r) \leq \mu(B^{-1})\ \nu(r),$$

where μ denotes any matrix norm compatible with ν. Similarly, if $\tilde{X}$ is an alleged inverse of B, and if ν is any multiplicative matrix norm we can calculate a bound for $\nu(\tilde{x} - B^{-1})$ in terms of the residual matrix $R = B\tilde{X} - I$;

$$\nu(\tilde{X} - B^{-1}) = \nu(B^{-1}R) \leq \nu(B^{-1})\ \nu(R).$$

For both problems we require a bound for $\nu(B^{-1})$. Such a bound is, in principle, easily constructed if we assume that B is similar to a diagonal matrix D:

$$B = SDS^{-1}.$$

For, assuming that ν is an axis-oriented norm and noting that $B^{-1} = SD^{-1}S^{-1}$ we have

(4.2-1) $$\nu(B^{-1}) \leq C\ (S)\ \lambda_{B^{-1}}.$$

If B were normal, then S could be taken unitary and the spectral condition number of S would be 1. In this case $\sigma(B^{-1}) \leq \lambda_{B^{-1}}$, and in view of Lemma 4,

$$\sigma(B^{-1}) = \lambda_{B^{-1}}.$$

Normal matrices are, of course, the only matrices unitarily similar to diagonal matrices. Estimates for non-normal matrices are not so easily derived. We see for instance that the bound (4.2-1), if at all applicable, requires the complete diagonalization of B.

If we let the function f^n be defined for all real $x \geq 0$ by

$$f^n(x) = x + x^2 + \cdots + x^n$$

we note that f and $x^{-1}f$ are monotonically increasing for $x > 0$, and that

$$\lim_{x \to 0^+} x^{-1} f^n(x) = 1.$$

With the notation of the preceeding section we have then:

THEOREM 10. *If* B *is non-singular and non-normal, and if* $\xi_1 = \Delta_1^{-1} \ell_1$, *then*

$$\sigma(B^{-1}) \leq \min \left[C_{\sigma}(N) \max_i \left\{ \frac{f^{n_i}(\xi_i)}{\xi_i} \Delta_i^{-1} \right\} \right]$$

where the minimum is taking over all ordered Schur forms.

Proof. As before we write $Q_{ii} = D_i + L_i$

$$\begin{aligned} Q_{ii}^{-1} &= (D_i + L_i)^{-1} \\ &= \left[D_i (I + D_i^{-1} L_i) \right]^{-1} \\ &= (I + D_i^{-1} L_i)^{-1} D_i^{-1} \end{aligned}$$

We cannot expand $(I + D_i^{-1} L_i)^{-1}$ according to the binomial expansion since D_i is not necessarily scalar. But if we expand without commuting it is still true that any term with more than $n_i - 1$ L_i's vanishes. Thus since

$$\sigma(D_i^{-1} L_i) \leq \sigma(D_i^{-1}) \sigma(L_i) = \Delta_i^{-1} \ell_i$$

we have, upon setting

$$\xi_i = \Delta_i^{-1} \ell_i,$$

$$\begin{aligned} \sigma(Q_{ii}^{-1}) &\leq (1 + \xi_i + \xi_i^2 + \cdots + \xi_i^{n_i - 1}) \Delta_i^{-1} \\ &= \frac{f^{n_i}(\xi_i)}{\xi_i} \Delta_i^{-1} \end{aligned}$$

In view of

$$\sigma(Q^{-1}) = \max_i \sigma(Q_{ii}^{-1})$$

and

$$\sigma(B^{-1}) = \sigma(A^{-1}),$$

$$\sigma(Q^{-1}) \leq \max_i \left\{ \frac{f^{n_i}(\xi_i)}{\xi_i} \Delta_i^{-i} \right\}$$

and

$$\sigma(B^{-1}) \leq C_\sigma(N) \max_i \left\{ \frac{f^{n_i}(\xi_i)}{\xi_i} \Delta_i^{-1} \right\}$$

for every ordered Schur form. The theorem follows.

CHAPTER 5

Spectral Variation and Eigenvalue Variation

Classical Results. Let the matrix $A = (a_{ij})$ have eigenvalues λ_i and let $B = (b_{ij})$ have eigenvalues μ_i, $i = 1,2,\ldots,n$. The quantity

$$s = s_A(B) = \max_{1 \le i \le n} \left\{ \min_{1 \le j \le n} |\mu_i - \lambda_j| \right\}$$

is called the spectral variation of B with respect to A. It is, in effect, the maximum distance from any eigenvalue of B to the spectrum of A. No one-to-one correspondence between eigenvalues is implied. However, the function v defined by

$$v = v(A,B) = \min_{\pi} \left\{ \max_{1 \le i \le n} |\lambda_i - \mu_{\pi(i)}| \right\}$$

where the minimum is taken with respect to all permutations of the set $(1,2,\ldots,n)$ and which is called the eigenvalue variation of A and B does imply a one-to-one correspondence. We have, of course, $v(A,B) = v(B,A)$ whereas $s_A(B) \neq s_B(A)$ in general. In addition

$$s_A(B) \le v(A,B)$$

for all matrices A and B.

One of the best available bounds for s and v for arbitrary matrices is given by Ostrowski [22]. (See also [24] p.192). Namely, if $M = \max_{1 \le i,j \le n} (|a_{ij}|, |b_{ij}|)$ and if the norm α is as defined in Chapter 1, then

$$s_A(B) \le (n + 2)\, M \left[\frac{\alpha(A - B)}{M} \right]^{1/n}$$

and

$$v(A, B) \le 2n(n + 2)\, M \left[\frac{\alpha(A - B)}{M} \right]^{1/n}.$$

That the exponent $1/n$ in these bounds cannot be improved in general may be seen by considering an example due to G. E. Forsythe (see [28], p.405). In special cases, however, improvements are possible.

If A is similar to a diagonal matrix D,

$$A = SDS^{-1}$$

and if ν is any axis-oriented lub norm, then Bauer and Fike [3] showed that

$$s_A(B) \le C_\nu(S)\, \nu(A - B). \tag{5.0-1}$$

If further, A is normal, S may be chosen unitary and we find for any norm ν majorizing the spectral norm

(5.0-2) $$S_A(B) \leq \nu(A - B).$$

If A and B are both normal and $\nu = \epsilon$ it follows from a result of Hoffman and Wielandt [15] that (5.0-2) is even valid for the eigenvalue variation:

(5.0-3) $$v(A,B) \leq \epsilon(A - B).$$

This result has been used frequently by Bargmann, Montogomery and von Neumann in [1] for A and B either real symmetric or hermitian.

A more recent result applicable to arbitrary matrices has been contributed by Henrici [14]. Because of the part that these estimates play in this chapter we shall develop the necessary notation. These estimates depend in particular upon a measure of non-normality which we define here.

If A is any matrix, we recall that (Mirsky, [19]) there exists a unitary matrix U and a triangular matrix T such that

$$A = UTU^*.$$

T, the Schur triangular form of A, is not uniquely determined for a given A. We put

$$T = D + M$$

where D denotes the diagonal matrix whose main diagonal coincides with that of T. It follows then that M is a strictly (lower)

triangular matrix.

If ν is a norm, the ν-departure from normality of A is defined by

$$\Delta_\nu(A) = \inf \nu(M)$$

where the infinum is taken with respect to all M that can appear in a Schur form. It follows that $\Delta_\nu(A) = 0$ if and only if A is normal.

Let the function $g = g(y)$ be defined for all real $y \geq 0$ as the (unique) non-negative solution of the equation

$$g + g^2 + \cdots + g^n = y.$$

The function g is the inverse of the function f defined in 4.2. For later use we note the relations

(5.0-4) $$\lim_{y \to 0^+} y^{-1} g(y) = 1$$

(5.0-5) $$n^{-1}y \leq g(y) \leq y, \qquad 0 \leq y \leq n$$

(5.0-6) $$g(n) = 1$$

(5.0-7) $$(n^{-1}y)^{1/n} \leq g(y) \leq y^{1/n}, \qquad y \geq n$$

(5.0-8) $$\lim_{y \to \infty} y^{-1/n} g(y) = 1.$$

Henrici's results may now be given.

THEOREM (Henrici). *Let* A *be a non-normal matrix, and let* $B - A \neq 0$. *If* ν *is any norm majorizing the spectral norm, and if*

$$\gamma = \frac{\Delta_\nu(A)}{\nu(B-A)}$$

then

$$S_A(B) \leq \frac{\gamma}{g(\gamma)}\,\nu(B-A) \tag{5.0-9}$$

and

$$\nu(A,B) \leq (2n-1)\,\frac{\gamma}{g(\gamma)}\,\nu(B-A). \tag{5.0-10}$$

(5.0-5), (5.0-6) and (5.0-7) may be used to render the bound (5.0-9) and (5.0-10) more explicit. For $\nu(B-A)$ bounded away from zero (5.0-4) shows that as $\Delta_\nu(A) \to 0$ the estimate (5.0-9) approaches (5.0-2). (5.0-8) shows that for a fixed non-normal A and for $B \to A$ the bound (5.0-9) is of the same order as (5.0-3).

It should be mentioned that Wielandt [29] had previously defined a measure of non-normality of a matrix. His measure is applicable only to matrices which are similar to a diagonal matrix, and requires the knowledge of a matrix effecting the diagonalization.

After deriving bounds for $S_A(B)$ and $\nu(A,B)$ using quasi-diagonal representations we shall make a comparison between these results and those of Henrici given above.

Section 5.1. Given an arbitrary matrix M, let us assume that a unitary U has been chosen such that $A = U^*MU$ is an ordered Schur form and that A has been transformed by N into $Q = \text{diag}\,(Q_{11}, Q_{22}, \ldots, Q_{kk})$ as indicated by Theorem 6. Let λ_{ij}, $j = 1,2,\ldots,n_i$ be the eigenvalues of Q_{ii}, $i = 1,2,\ldots,k$.

Let B be an arbitrary matrix. We have from above:

$$N^{-1}AN = Q = \text{diag}\,(Q_{11}, Q_{22}, \ldots, Q_{kk})$$

$$Q_{ii} = D_i + L_i,$$

D_i being the diagonal matrix whose diagonal elements coincide with those of Q_{ii}.

Let

$$N^{-1}BN = B_1$$

$$E = B - A, \qquad N^{-1}EN = F.$$

Then

$$B_1 = F + Q.$$

Let μ be an arbitrary but fixed eigenvalue of B (and hence of B_1) which is not an eigenvalue of A. Then $(Q - \mu I)^{-1}$ exists and

$$0 = \det (B_1 - \mu I) = \det [Q - \mu I + F]$$

$$= \det [Q - \mu I] \det [I + (Q - \mu I)^{-1} F] .$$

Thus

$$\det [I + (Q - \mu I)^{-1} F] = 0$$

and -1 is an eigenvalue of $(Q - \mu I)^{-1}F$. By Lemma 4

$$\sigma\left[(Q - uI)^{-1} F\right] \geq 1.$$

But

$$\sigma(F) \leq C_{\sigma}(N)\ \sigma(E)$$

so

$$\sigma\left[(Q - \mu I)^{-1}\right] \geq \frac{1}{C_{\sigma}(N)\ \sigma(E)} . \tag{5.1-1}$$

$(D_i - \mu I)$ is non-singular since D contains only eigenvalues of A and consequently

$$(Q_{ii} - \mu I)^{-1} = (D_i + L_i - \mu I)^{-1} \tag{5.1-2}$$

$$= \left\{(D_i - \mu I)\left[I + (D_i - \mu I)^{-1} L_i\right]\right\}^{-1}$$

$$= \left[I + (D_i - \mu I)^{-1} L_i\right]^{-1} (D_i - \mu I)^{-1} .$$

Now

$$(5.1\text{–}3) \qquad \left[I + (D_i - \mu I)^{-1} L_i\right]^{-1} = I - (D_i - \mu I)^{-1} L_i$$

$$+ \left[(D_i - \mu I)^{-1} L_i\right]^2 - \cdots (-1)^{n_i - 1} \left[(D_i - \mu I)^{-1} L_i\right]^{n_i - 1}.$$

The last sum extends to at most the $n_i - 1$ power since $\left[(D_i - \mu I)^{-1} L_i\right]^r = 0, \quad r \geq n_i$. The validity of this statement can be verified by noting that $(D_i - \mu I)^{-1} L_i$ has the same (i.e. lower triangular) form as L_i, and that $L_i^r = 0, \quad r \geq n_i$.

Let $p_i = \sigma\left[(D_i - \mu I)^{-1}\right]$

$$\sigma(L_i) = \ell_i$$

$$\sigma(E) = e$$

$$C_\sigma(N) = x.$$

Note that

$$p_i = \max_{1 \leq j \leq n_i} \left\{\left|\lambda_{ij} - \mu\right|^{-1}\right\} = \frac{1}{\min\limits_{1 \leq j \leq n_i} |\lambda_{ij} - \mu|}.$$

From (5.1–2) and (5.1–3)

$$\sigma\left\{\left[I + (D_i - \mu I)^{-1} L_i\right]^{-1}\right\} \leq 1 + p_i \ell_i + p_i^2 \ell_i^2 + \cdots + p_i^{n_i - 1} \ell_i^{n_i - 1}$$

and

$$(5.1\text{–}4) \qquad \sigma\left[(Q_{ii} - \mu I)^{-1}\right] \le p_i + p_i^2 \ell_i + \cdots + p_i^{n_i} \ell_i^{n_i - 1}$$

$$= \frac{f^{n_i}(p_i \ell_i)}{\ell_i} .$$

But

$$(5.1\text{–}5) \qquad \sigma\left[(Q - \mu I)^{-1}\right] = \max_{1 \le i \le k} \sigma\left[(Q_{ii} - \mu I)^{-1}\right].$$

Thus from (5.1–1), (5.1–4) and (5.1–5)

$$\frac{1}{xe} \le \max_{1 \le i \le k} \sigma\left[(Q_{ii} - \mu I)^{-1}\right] \le \max_{1 \le i \le k} \frac{f^{n_i}(p_i \ell_i)}{\ell_i} .$$

For some index $i = i_0$ say

$$\frac{1}{xe} \le \frac{f^{n_{i_0}}(p_{i_0} \ell_{i_0})}{\ell_{i_0}}$$

$$\frac{\ell_{i_0}}{xe} \le f^{n_{i_0}}(p_{i_0} \ell_{i_0})$$

and from the definition of $g^{n_{i_0}}$

$$g^{n_{i_0}}\left(\frac{\ell_{i_0}}{xe}\right) \le p_{i_0} \ell_{i_0}$$

$$\frac{1}{p_{i_0}} \leq \frac{\ell_{i_0}}{n_{i_0}\left(\frac{\ell_{i_0}}{xe}\right)}.$$

Then

$$\frac{1}{p_{i_0}} \leq \max_{1 \leq i \leq k} \frac{\ell_i}{g^{n_i}\left(\frac{\ell_i}{xe}\right)}$$

and

$$\min_{1 \leq j \leq n_{i_0}} |\lambda_{i_0 j} - \mu| \leq \max_{1 \leq i \leq k} \frac{\ell_i}{g^{n_i}\left(\frac{\ell_i}{xe}\right)}. \tag{5.1-6}$$

Thus for any arbitrary eigenvalue $\mu = \mu_r$ of B it is possible to find some eigenvalue of A such that (5.1-6) holds.

From (5.1-6),

$$\min_{\substack{1 \leq i \leq k \\ 1 \leq j \leq n_i}} |\lambda_{ij} - \mu_r| \leq \max_{1 \leq i \leq k} \frac{\ell_i}{g^{n_i}\left(\frac{\ell_i}{xe}\right)}$$

for $r = 1, 2, \ldots, n$ and

$$S_A(B) = \max_{1 \leq r \leq n} \min_{i,j} |\lambda_{ij} - \mu_r| \leq \max_{1 \leq i \leq k} \frac{\ell_i}{g^{n_i}\left(\frac{\ell_i}{xe}\right)}.$$

THEOREM 11. Let A be a non-normal matrix, and let $B - A \neq 0$.

If

$$y_i = \frac{\sigma(L_i)}{C_\sigma(N)\ \sigma(B - A)}$$

then

$$S_A(B) \leq \left\{ \max_{1 \leq i \leq k} \frac{y_i}{g^{n_i}(y_i)} \right\} C_\sigma(N)\ \sigma(B - A).$$

Here L_i and n_i are as defined earlier in this section.

If ν is any norm majorizing σ, we have since g is non-negative and monotone increasing

$$\frac{\sigma(L_i)}{g^{n_i}\left[\frac{\sigma(L_i)}{C_\sigma(N)\ \sigma(B - A)}\right]} \leq \frac{\sigma(M_i)}{g^{n_i}\left[\frac{\sigma(M_i)}{\nu(B - A)C_\nu(N)}\right]}$$

and consequently

$$S_A(B) \leq \max_{1 \leq i \leq k} \frac{y_i}{g^{n_i}(y_i)}\ C_\nu(N)\ \nu(B - A) \tag{5.1-7}$$

where

$$y_i = \frac{\sigma(M_i)}{C_\nu(N)\ \nu(B - A)}.$$

Let now $0 < y_1 < y_2$ and define $x_i = g(y_i)$, $i = 1,2$. From the monotonicity of $x^{-1}(f(x))$ it follows that

$$\frac{y_1}{g(y_1)} = \frac{f(x_1)}{x_1} < \frac{f(x_2)}{x_2} = \frac{y_2}{g(y_2)} .$$

Thus the function $y[g(y)]^{-1}$ is monotonically increasing, and we have from (5.1–7) replacing $\sigma(L_i)$ by $\nu(L_i)$;

COROLLARY. *If* A *is non-normal and* $B - A \neq 0$ *we have for any norm* ν *majorizing* σ:

$$(5.1\text{–}8) \qquad S_A(B) \leq \left\{ \max_{1 \leq i \leq k} \frac{y_i}{n_i g(y_i)} \right\} C_\nu(N)\ \nu(B - A)$$

where

$$y_i = \frac{\nu(L_i)}{C_\nu(N)\ \nu(B - A)} .$$

Section 5.2. Comparison of bounds and related results.

Our object here shall be to compare the results given by Henrici [14]:

$$(5.0\text{-}9) \qquad S_A(B) \leq \frac{y}{g^n(y)}\, \nu(B - A), \qquad y = \frac{\Delta_\nu(A)}{\nu(B - A)}$$

with that derived earlier (Corollary to Theorem 11).

$$(5.1\text{-}8) \quad S_A(B) \leq \max_{1 \leq i \leq k} \frac{y_i}{g^{n_1}(y_i)}\, C_\nu(N)\, \nu(B - A), \quad y_i = \frac{\nu(L_i)}{C_\nu(N)\, \nu(B-A)}.$$

For this purpose, let

$$z_i = \frac{\nu(L_i)}{\Delta_\nu(A)}\, y$$

and note that

$$y_i = \frac{z_i}{C_\nu(N)}.$$

For those norms ν such that $\nu(L_i) < \Delta_\nu(A)$, $z_i < y$ and we have by the monotonicity of $y[g(y)]^{-1}$

$$(5.2\text{-}1) \qquad \frac{z_i}{g^n(z_i)} < \frac{y}{g^n(y)}.$$

Let $K = C_\nu(N)$. Then for those values of z_i such that

$$g^{n_i}\left(\frac{z_i}{K}\right) > g^n(z_i) \tag{5.2-2}$$

we have from (5.2-1)

$$\frac{z_i}{g^{n_i}\left(\frac{z_i}{K}\right)} < \frac{y}{g^n(y)} .$$

Writing this inequality in terms of y_i

$$\frac{y_i C_{y_i}(N)}{g^{n_i}(y_i)} < \frac{y}{g^n(y)} .$$

Thus, for those values of z_i such that $z_i < y$ and such that (5.2-2) holds, the estimate (5.1-8) is an improvement over (5.0-9) given by Hentici.

We shall then be interested in determining conditions on z in terms of n, n_i, K such that

$$g^{n_i}\left(\frac{z}{K}\right) > g^n(z) .$$

For this purpose let us introduce the function $h(z)$ defined by:

$$h(z) = g^n(z) - g^{n_i}\left(\frac{z}{K}\right)$$

and determine those values of z such that $h(z) < 0$. We may, of course, assume that $K > 1$, for $K = 1$ implies $h(z) < 0$ for

positive values of z.

We begin by investigating the positive zeros of $h(z)$.

Define

$$p(x) = x^n + x^{n-1} + \cdots + x^{n_1+1} + (1-K)x^{n_1} + (1-K)x^{n_1-1} + \cdots + (1-K)x.$$

LEMMA 16.

$$h(z) = 0 \quad \underline{\text{if and only if}} \quad p(x) = 0,$$

where

$$x = g^n(z) > 0 \quad [z = f^n(x)].$$

Proof. Suppose $h(z) = 0$ for some $z > 0$. Then $g^n(z) = g^{n_1}(z/K)$.

Letting x be this common value we have

$$z = x + x^2 + \cdots + x^n$$
$$\frac{z}{K} = x + x^2 + \cdots + x^{n_1}.$$

Then

$$p(x) = x^n + x^{n-1} + \cdots + x - K(x^{n_1} + x^{n_1-1} + \cdots + x) = 0.$$

On the other hand if $p(x) = 0$,

$$x^n + x^{n-1} + \cdots + x = K(x^{n_i} + x^{n_i-1} + \cdots + x).$$

Setting $z = f^n(x)$ we have

$$\frac{z}{K} = f^{n_i}(x)$$

and consequently

$$g^n(z) = g^{n_i}\left(\frac{z}{K}\right)$$

proving our assumption.

We have shown that the positive (non-negative) zeros of $h(z)$ are in a one-to-one correspondence with those of $p(x)$.

By Descarte's rule of signs $p(x)$ can have at most one positive zero, say x_0. By the above lemma, $h(z) = 0$ for at most positive value of z, namely $z_0 = f^n(x_0)$.

Recalling that

$$h(z) < 0 \text{ if and only if } g^{n_i}\left(\frac{z}{K}\right) > g^n(z)$$

we are leading to the following lemmas.

LEMMA 17. *If* $g^{n_i}(z/K) > g^n(z)$ *for some* z, *then* $p(x) > 0$ *where* $x = g^{n_i}(z/K)$ *and* $p(w) > 0$ *where* $w = g^n(z)$.

Proof. Letting $x = g^{n_i}(z/K)$, $w = g^n(z)$ we have

$$x^{n_i} + x^{n_i-1} + \cdots + x = \frac{z}{K}$$

$$w^{n} + w^{n-1} + \cdots + w = z$$

and thus

$$w^{n} + w^{n-1} + \cdots + w = K(x^{n_i} + x^{n_i-1} + \cdots + x).$$

By hypothesuse, $x > w$ and

$$f^{n}(x) > f^{n}(w) = Kf^{n_i}(x) > Kf^{n_i}(w).$$

Therefore

$$p(x) = f^{n}(x) - Kf^{n_i}(x) > 0$$

$$p(w) = f^{n}(w) - Kf^{n_i}(w) > 0.$$

LEMMA 18. *If* $p(x) > 0$ *for some* x *then* $f^{n_i}(z/K) > g^{n}(z)$, *where* $z = f^{n}(x)$.

Proof. $p(x) > 0$ implies that

$$x^{n} + x^{n-1} + \cdots + x > K(x^{n_i} + x^{n_i-1} + \cdots + x)$$

or

$$x^{n} + x^{n-1} + \cdots + x = K(x^{n_i} + x^{n_i-1} + \cdots + x) + Kq$$

for some $q > 0$.

Let

$$w = g^{n_i}\left[f^{n_i}(x) + q\right]$$

so that

$$f^{n_i}(w) = f^{n_i}(x) + q.$$

Then $w > x$, and

$$x^n + x^{n-1} + \cdots + x = K(w^{n_i} + w^{n_i-1} + \cdots + w).$$

Since $z = f^n(x)$, $z/K = f^{n_i}(w)$ and $f^{n_i}(z/K) = w > x = g^n(z)$. Combining these two lemmas we see that

$h(z) < 0$ *if and only if* $p(x) > 0$, $z = f^n(x)$ $[x = g^n(z)]$.

Since $p(x) > 0$ for large x, and since $p(x)$ has at most one positive root, say x_0, $p(x) > 0$ for all $x > x_0$. Consequently $h(z) < 0$ for all $z > z_0 = f^n(x_0)$. If $p(x)$ has no positive zero, $p(x) > 0$ for all $x > 0$ and $h(z) < 0$ for all $z > 0$.

The determination of positive zeros of $p(x)$ is, in general, a difficult problem. We shall instead determine portions of the positive x-axis for which $p(x) > 0$. It is of course only necessary to find any point $x_0 > 0$ for which $p(x_0) > 0$. For then $p(x) > 0$, $x \geq x_0$.

There are three cases to consider.

Case 1. $Kn_i = n$.

In this case $p(1) = 0$ and

$p(x) > 0$ for $x > 1$;

$h(z) < 0$ for $z > n$.

Case 2. $Kn_i > n$.

If $x \neq 1$, we may rewrite $p(x)$ as

$$p(x) = \frac{x^{n+1} - 1}{x - 1} - K \frac{x^{n_i+1} - 1}{x - 1} .$$

Then

$$r(x) = (x - 1)\, p(x)$$
$$= x^{n+1} - K x^{n_i+1} + (K - 1)$$

has the same sign as $p(x)$ for $x > 1$ and opposite sign for $x < 1$. But

$$r\left(K^{\frac{1}{n-n_i}}\right) = K^{\frac{n+1}{n-n_i}} - K^{1+\frac{n_i+1}{n-n_i}} + (K - 1)$$

$$= (K - 1) > 0.$$

Thus $p(K^{\frac{1}{n-n_i}}) > 0$ since $K^{\frac{1}{n-n_i}} > 1$. Hence

$$p(x) > 0$$

at least for $x \geq K^{\frac{1}{n-n_i}}$ and $h(z) < 0$ at least for $z \geq f^n(K^{\frac{1}{n-n_i}})$.

Case 3. $Kn_i < n$.

$$p(1) = n - Kn_i > 0$$

$$p(\kappa) > 0 \quad \text{for} \quad \kappa \geq 1$$

$$h(z) < 0 \quad \text{for} \quad z \geq f^n(1) = n.$$

THEOREM 12. For all norms ν, such that $\nu(L_i) < \Delta_\nu(A)$, $i = 1,2,\ldots,k$

1. $g^{n_i}\left(\frac{z_i}{K}\right) > g^n(z_i)$ for $z_i > n$ if $K n_i = n$

2. $g^{n_i}\left(\frac{z_i}{K}\right) > g^n(z_i)$ for $z_i \geq f^n\left(K^{\frac{1}{n-n_i}}\right)$ if $K n_i > n$

3. $g^{n_i}\left(\frac{z_i}{K}\right) > g^n(z_i)$ for $z_i \geq n$ if $K n_i < n$.

Since $y = \frac{\Delta_\nu(A)}{\nu(L_i)} z_i$ we have

COROLLARY. For all norms ν such that

$$\nu(L_i) < \Delta_\nu(A), \qquad i = 1,2,\ldots,k$$

$$\frac{y_i C_\nu(N)}{g^{n_i}(y_i)} < \frac{y}{g^n(y)}$$

for

1. $y > n \frac{\Delta_\nu(A)}{\nu(L_i)}$ if $Kn_i = n$

2. $y \geq \frac{\Delta_{\mathcal{V}}(A)}{\mathcal{V}(L_i)} \; f^n\left(K^{\frac{1}{n-n_i}}\right)$ if $Kn_i > n$

3. $y \geq n \frac{\Delta_{\mathcal{V}}(A)}{\mathcal{V}(L_i)}$ if $Kn_i < n$.

Therefore the estimate (5.1-8) represents an improvement over (5.0-9) if for each i

1. $y > n \max_{1 \leq i \leq k} \frac{\Delta_{\mathcal{V}}(A)}{\mathcal{V}(L_i)}$ for $Kn_i = n$

2. $y \geq f^n\left(K^{\frac{1}{n-n_i}}\right) \max_{1 \leq i \leq k} \frac{\Delta_{\mathcal{V}}(A)}{\mathcal{V}(L_i)}$ for $Kn_i > n$

3. $y \geq n \max_{1 \leq i \leq k} \frac{\Delta_{\mathcal{V}}(A)}{\mathcal{V}(L_i)}$ for $Kn_i < n$.

The intersection of the regions determined for each i above yields an interval of the y axis for which our estimate is preferable.

The above analysis is valid for every ordered Schur form A. Therefore we have:

THEOREM 13. Under the hypotheses of Theorem 11,

$$S_A(B) \leq \min\left\{\left[\max_{1 \leq i \leq k} \frac{y_i}{g^{n_i}(y_i)}\right] C_{\mathcal{V}}(N)\, \mathcal{V}(B - A)\right\}$$

where

$$y_i = \frac{\nu(L_i)}{C_\nu(N)\,\nu(B - A)}$$

and the minimum is taken with respect to all ordered Schur forms.

From the fact that $B - A = U(U^*BU - M)U^*$ and the unitary invariance of σ, $\sigma(B - A) = \sigma(U^*BU - M)$ and we may rewrite the above theorem as follows:

COROLLARY. For non-normal M with $M - B \neq 0$ we have for any norm ν dominating σ

$$S_M(B) \leq \min\left\{\left[\max_{1 \leq i \leq k} \frac{y_i}{g^{n_i}(y_i)}\right] C_\nu(N)\,\nu(U^*BU - M)\right\} \tag{5.2-3}$$

where

$$y_i = \frac{\nu(L_i)}{C_\nu(N)\,\nu(U^*BU - M)}$$

and the minimum is taken with respect to all U occurring in an ordered Schur form of M.

Related results on $\nu(M,B)$

For given matrices M, B satisfying the hypotheses of Theorem 11 let δ represent the quantity on the right hand side of (5.2-3). The statement of the above corollary may then be interpreted geometrically by saying that the spectrum of B is contained in the union I_δ of the disks

$$D_i = \{\lambda : |\lambda - \lambda_i| \leq \delta\} \qquad i = 1,2,\ldots,k.$$

Since $\delta \longrightarrow 0$ monotonically as $U^*BU \longrightarrow M$, or alternately, as $B \longrightarrow A$ we may conclude by a well known continuity argument (see e.g. [22]) that each component of L_δ contains as many eigenvalues of B as of M. From this fact we can obtain, again using a well-known argument (see especially the translator's note in [23]) the following result:

THEOREM 14. *For non-normal* M *with* $M - B \neq 0$ *we have for any norm* ν *dominating* σ:

$$v(M,B) \leq (2k - 1) \min \left\{ \left[\max_{1 \leq i \leq k} \frac{y_i}{g^{n_i}(y_i)} \right] C_\nu(N)\, \nu(U^*BU - M) \right\}$$

where

$$y_i = \frac{\nu(L_i)}{C_\nu(N)\, \nu(U^*BU - M)}$$

and where the minimum is taken with respect to all U *occurring in an ordered Schur form of* M.

BIBLIOGRAPHY

[1] Bargmann, V., Montgomery, D., and von Neumann, J., Solution of linear systems of high order. Report prepared under contract NORD 9596 with the Bureau of Ordnance, Navy Department, Institute for Advanced Study, 1946.

[2] Bauer, F. L. On the definition of condition numbers and their relation to closed methods for solving linear systems. Proceedings of the International Conference on Information Processing Unesco, Paris 1960. pp. 109–110.

[3] Bauer, F. L., and Fike, C. T., Norms and exclusion theorems. Numerische Math. 2, (1960). 137–141.

[4] Bellman, R., Introduction to Matrix Analysis. McGraw-Hill, New York, 1960.

[5] Cordes, H. O., Über die Spektralzerlegung von hypermaximalen Operatoren, die durch Separation der Variablen zerfallen, I, Math. Ann. 128, (1954), 257–289.

[6] _____, Über die Spektralzerlegung von hypermaximalen Operatoren, die durch Separation der Variablen zerfallen, II, Math. Ann. 128, (1955), 373–411.

[7] Dahlquist, G., Stability and error bounds in the numerical integration of ordinary differential equations. Trans. Roy. Inst. Technol. Stockholm, Nr. 130, 1959.

[8] Forsythe, G. E., Singularity and near singularity in numerical analysis. Amer. Math. Monthly 65, (1958), 229–240.

[9] Gautschi, W., The asymptotic behavior of powers of matrices. Duke Math. J. 20, (1953), 127–140.

[10] _____, The asymptotic behavior of powers of matrices. II. Duke Math. J. 20, (1953), 375–379.

[11] Givens, W., Elementary divisors and some properties of the Lyapunov Mapping $X \to AX + XA^*$. Argonne National Laboratory report 6456 (1961) (pp. 73–75 in particular).

[12] Hahn, W., Theorie und Anwendung der direkten Methode von Ljapunov. Ergebnisse der Mathematik und Ihrer Grenzgebiete, p. 25 (1959).

[13] Heinz, E., Beiträge zur Störungstheorie der Spektralzerlegung. Math. Ann. 123, (1951), 415–438.

[14] Henrici, P. K., Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices. Numerische Math. 4, (1962), 24–40.

[15] Hoffman, A. J., and Wielandt, H. W. The variation of the spectrum of a normal matrix. Duke Math. J. 20, (1953), 37–39.

[16] Householder, A. S., The approximate solution of matrix problems. J. Assoc. Comput. Mach. 5, (1958), 204–243.

[17] Kato, T., Estimation of iterated matrices, with application to the von Neumann condition. Numerische Math. 2, (1960), 22–29.

[18] Lax, P. D. and Richtmeyer, R. D., Survey of the stability of linear finite difference equations. Comm. Pure Appl. Math. 9, (1956), 267–293.

[19] Mirsky, L., An introduction to linear algebra. Oxford University Press, Oxford, 1955.

[20] Ostrowski, A., On some metrical properties of operator matrices and matrices partitioned into blocks. University of Wisconsin, Mathematics Research Center Report 138 (1960).

[21] _____, Über Normen von Matrizen. Math. Z. 63, (1955), 2–18.

[22] _____, Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen. Jahresber. Deutsch. Math. Verein. 60, Abt. 1, (1957), 40–42.

[23] _____, On the continuity of characteristic roots in their dependence on the matrix elements [translation of [22]]. Translated from the German and annotated by G. E. Bump. Technical Report No. 2, Applied Mathematics and Statistics Laboratories, Stanford University, 1959.

[24] _____, Solution of equations and systems of equations. Academic Press, New York and London, 1960.

[25] Rosenblum, M., On the operator equation $BX - XA = Q$. Duke 23, (1956), 263–269.

[26] Rutherford, D. E., On the solution of the matrix equation $AX + XB = C$. Akademie van Wetenschappen, Amsterdam Proc., Afdeeling Natuurkunde 35, Part I, (1932), 54–59.

[27] Stone, B. J., Best possible ratios of certain matrix norms. Numerische Math. 4, (1962), 114-116.

[28] White, P. A., The computation of eigenvalues and eigenvectors of a matrix. J. Soc. Indust. Appl. Math. 6, (1958), 393-437.

[29] Wielandt, H., Inclusion theorems for eigenvalues. Nat. Bur. Standards Appl. Math. Ser. 29, (1951), 75-78.

APPENDIX

The Matrix Equation $-AX + XB = C$ and Related Results

The problem of finding a solution to the matrix equation $-AX + XB = C$ where A, B are square matrices of arbitrary order be denoted as problem (A). It is of importance in the development of a canonical form, which in turn is applied to the solution of problems (ii), (iii) and (iv) as set forth in the Introduction.

Let $\mathcal{B}$ be a Banach algebra, with elements $A, B, Q, \ldots$. T will be an operator on $\mathcal{B}$ such that

$$T(X) = -AX + XB \quad \text{for every} \quad X \in \mathcal{B}.$$

The following results are to be found in the literature.

Result 1. [Rutherford, [26]]

Let $\mathcal{B}$ be the algebra of $n \times n$ matrices. If the characteristic roots of A are distinct from the characteristic roots of B, then T^{-1} exists and is bounded.

The proof, though constructive, depends upon a complete knowledge of the Jordan Canonical form.

Result 2. [Heinz, [13]]

Let $\mathcal{B}$ be the space of bounded linear operators on a Hilbert space, $\mathcal{H}$, with inner product $(\cdot,\cdot)$. If there exist real numbers a and b such that $a > b$, $B + B^* \leq b$, $A + A^* \geq a$,

then T^{-1} exists as a bounded linear operator and has the

representation

$$T^{-1}(Q) = -\int_0^\infty e^{-At} Q e^{Bt} dt,$$

where by $A \leq \alpha$ we mean $(u, Au) \leq \alpha$ for all $u \in \mathcal{H}$.

Notable extensions of Result 2 are given by Rosenblum [25] and Cordes [5], [6].

Result 3.

Givens [11] gives a formal solution of $AX + XA^T = Y$ in terms of adjoints, which although not done in his paper, is immediately extendable to the problem $AX + XB = Y$. The proof and subsequent simplicity of the representation depend upon the assumption of simplicity of the roots of A (or B), a severe restriction for our purposes. These results generalize those of Hahn [12] for the case $Y = I$.

We shall see that an integral representation of the solution of $-AX + XB = C$ similar to that given by Result 2 is valid under assumptions similar to those in Result 1 but without the restriction that A, B and C be square matrices of order n.

THEOREM A-1. A necessary and sufficient condition that problem (A) have a solution for all C is that $-\lambda_i + \mu_j \neq 0$ where λ_i are the eigenvalues of A and μ_j the eigenvalues of B; i.e., if and only if the eigenvalues of A differ from those of B. If a solution exists it is unique.

Proof Let $A \otimes B$ denote the Kronecker product of arbitrary

matrices A and B. That is, $A \otimes B$ is a matrix whose general element is $a_{ij}B$. The eigenvalues of $A \otimes B$ are all the possible products $\lambda_i \mu_j$ where λ_i is any eigenvalue of A and μ_j is any eigenvalue of B. [Bellman, [4]].

If we consider $-AX + XB = C$ as a system of linear equations in the unknowns x_{ij} the coefficient matrix is $-A \otimes I + I \otimes B^T$, the roots of which are $-\lambda_i + \mu_j$. By assumption $-\lambda_i + \mu_j \neq 0$. Therefore a (unique) solution exists.

Result 4. [Bellman [4], p.175]

If the expression

$$X = -\int_0^\infty e^{-At} C e^{Bt} \, dt$$

exists for all C, it represents the unique solution of

$$-AX + XB = C.$$

The existence of the integral implies that $\lim_{t \to \infty} z(t) = 0$ where

$$z(t) = e^{-At} C e^{Bt}.$$

We shall examine the form of $z(t)$ in detail. For any square scalar matrix G, let J denote its Jordan Canonical form. Thus there exists a nonsingular constant matrix T such that

$$G = TJT^{-1}.$$

J has the form

$$J = \begin{pmatrix} J_0 & 0 & 0 & \dots & 0 \\ 0 & J_1 & 0 & \dots & 0 \\ 0 & 0 & J_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & \dots & J_s \end{pmatrix}$$

where J_0 is a diagonal matrix with diagonal entries $\eta_1, \eta_2, \dots, \eta_q$ and

$$J_i = \begin{pmatrix} \eta_{q+i} & 1 & 0 & \dots & 0 & 0 \\ 0 & \eta_{q+i} & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & \eta_{q+i} & 1 \\ 0 & 0 & 0 & \dots & 0 & \eta_{q+i} \end{pmatrix} (i = 1, 2, \dots, s)$$

$$= \eta_{q+i} I_{r_i} + Z_i$$

where

$$Z_i = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

is of order r_i.

It follows that

$$e^{tJ} = \begin{pmatrix} e^{tJ_0} & 0 & \cdots & 0 \\ 0 & e^{tJ_1} & \cdots & 0 \\ \cdot & \cdot & \cdots & 0 \\ 0 & 0 & \cdots & e^{tJ_s} \end{pmatrix}$$

and

$$e^{tJ_0} = \begin{pmatrix} e^{t\eta_1} & 0 & \cdots & 0 \\ 0 & e^{t\eta_2} & \cdots & 0 \\ \cdot & \cdot & \cdots & 0 \\ 0 & 0 & \cdots & e^{t\eta_q} \end{pmatrix}$$

Since $J_i = \lambda_{q+i} I_{r_i} + z_i$ and from the fact that $\eta_{q+i} I_{r_i}$ commutes with z_i we have

$$e^{tJ_i} = e^{\eta_{q+i} t} e^{z_i t}.$$

Thus

$$e^{tJ_i} = e^{\eta_{q+i} t} \begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{r_i-1}}{(r_i-1)!} \\ 0 & 1 & t & \cdots & \frac{t^{r_i-2}}{(r_i-2)!} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where J_i is an $r_i \times r_i$ matrix.

Hence if J_A, J_B are, respectively, Jordan Canonical forms for A and B and T, V are nonsingular matrices such that

$$A = TJ_AT^{-1}, \quad B = VJ_BV^{-1}$$

it follows that

$$e^{-At} = T e^{-tJ_A} T^{-1}, \quad e^{Bt} = V e^{tJ_B} V^{-1}$$

and

$$z(t) = T e^{-tJ_A} T^{-1} CV e^{tJ_B} V^{-1}.$$

Thus every element of $z(t)$ is a linear combination of terms of the form

$$e^{(-\lambda_i + \mu_j)t} p_k(t)$$

where $p_k(t)$ is a polynomial of degree not exceeding $n + m - 2$ if A is of order n and B is of order m.

It is clear then that if $\operatorname{Re} \lambda_i > \operatorname{Re} \lambda_j$ $\lim_{t \to \infty} z(t) = 0$. Moreover it is clear in this case that the integral exists for all C since each element of $z(t)$ is integrable. We have then the following lemma.

LEMMA A. *A sufficient condition that*

$$X = -\int_0^\infty e^{-At} Ce^{Bt} dt$$

be a solution of the matrix equation

$$-AX + XB = C$$

is that $\operatorname{Re} \lambda_i > \operatorname{Re} \mu_j$ for all λ_i and μ_j, eigenvalues of A and B respectively. The solution, if it exists, is unique.

We shall now show that it is possible to give an integral formula for X similar to Result 2 which will be valid if $\operatorname{Re} \lambda_i \succ \operatorname{Re} \mu_j$ for all λ_i, μ_j.

$\operatorname{Re} \lambda_i \succ \operatorname{Re} \mu_j$ implies that either

(1) $\operatorname{Re} \lambda_i > \operatorname{Re} \lambda_j$

or

(2) $\operatorname{Re} \lambda_i = \operatorname{Re} \mu_j$ and $\operatorname{Im} \lambda_i > \operatorname{Im} \mu_j$.

We have already disposed of (1) in the preceeding lemma and must now consider the case where (2) holds for some or all of the roots of A and B.

Instead, however, of trying to find a direct solution of

(A-1) $$-AX + XB = C,$$

we shall solve the system

(A-2) $$-e^{i\theta} AX + Xe^{i\theta} B = Ce^{i\theta},$$

where θ is a real number to be determined

Clearly any solution of (A-2) will be a solution of (A-1) and the unique solution will be given by

$$(A\text{-}3) \qquad X = -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} At\right] C \exp\left[e^{i\theta} Bt\right] dt$$

provided the integral exists.

The idea is to demonstrate values of θ such that $\operatorname{Re} e^{i\theta}\lambda_i > \operatorname{Re} e^{i\theta}\mu_j$ for all λ_i, μ_j, and to apply the preceeding lemma yielding the representation given by (A-3).

For this, let an arbitrary root of A be given by $\lambda_i = x_i + iy_i$; and that of B be $\mu_j = u_j + iv_j$.

The roots of $-e^{i\theta} A$ are $(-x_i \cos\theta + y_i \sin\theta) + i(-y_i \cos\theta - x_i \sin\theta)$ and those of $e^{i\theta} B$ are $(u_j \cos\theta - v_j \sin\theta) + i(v_j \cos\theta + u_j \sin\theta)$.

Letting

$$h_{ij}(\theta) = \left[(u_j - x_i)\cos\theta + (y_i - v_j)\sin\theta\right] + i\left[(v_j - y_i)\cos\theta + (u_j - x_i)\sin\theta\right]$$

we have that every term of

$$z_\theta(t) = e^{i\theta} \exp\left[-e^{i\theta} At\right] C \exp\left[e^{i\theta} Bt\right]$$

consists of linear combinations of terms of the form $e^{h_{ij}(\theta)} p_k(t)$, where $p_k(t)$ are polynomials in t of finite degree.

A sufficient condition that $\lim_{t\to\infty} z_\theta(t) = 0$, and indeed that the integral (A-3) exist is that

$$(A\text{-}4) \qquad \operatorname{Re} h_{ij}(\theta) < 0 \quad \text{for all } i, j.$$

We shall show that it is always possible to choose a θ such that (A–4) holds. Indeed θ can be chosen from a non–degenerate interval.

Let

(A–5) $$\alpha_{ij} = -x_i + u_j$$

(A–6) $$\beta_{ij} = y_i - v_j.$$

Then

$$\text{Re } h_{ij}(\theta) = \alpha_{ij} \cos \theta + \beta_{ij} \sin \theta.$$

We have, by the lexicographic ordering:

$$\alpha_{ij} \leq 0 \quad \text{and} \quad \alpha_{ij} = 0 \quad \text{implies} \quad \beta_{ij} > 0.$$

If all $\alpha_{ij} < 0$ it is sufficient to choose $\theta = 0$.

If $\alpha_{ij} = 0$ then $\text{Re } h_{ij}(\theta) < 0$ iff $\sin \theta < 0$, i.e. $\pi < \theta < 2\pi$.

If some $\alpha_{ij} < 0$, for $\text{Re } h_{ij}(\theta) < 0$ the following relationships must be true:

$$\alpha_{ij} \cos \theta + \beta_{ij} \sin \theta < 0$$

$$\beta_{ij} \sin \theta < -\alpha_{ij} \cos \theta.$$

If, in addition, $\pi < \theta < 2\pi$,

$$\frac{\sin \theta}{-\alpha_{ij}} < 0$$

and

$$\frac{\beta_{ij}}{-\alpha_{ij}} > \cot \theta .$$

(A-7) Let $\gamma_{ij} = \beta_{ij}/-\alpha_{ij}$.

If

(A-8) $$\gamma = \min_{i,j} \gamma_{ij}$$

$\cot \theta < \gamma$ implies that $\mathrm{Re}\, h_{ij}(\theta) < 0$. $\gamma \neq -\infty$ since this would mean that for some i, j, $\alpha_{ij} = 0$, $\beta_{ij} < 0$ which contradicts the assumed lexicographic ordering. Since $\cot \theta$ assume all value between $-\infty$ and $+\infty$ in the interval $\pi < \theta < 2\pi$, it is sufficient to choose θ such that

(A-9) $$\pi < \cot^{-1} \gamma < \theta < 2\pi .$$

LEMMA B. *If γ is defined by (A-8) and if θ is such that (A-9) holds then*

$$X = -\int_0^\infty e^{i\theta} \exp\left[-e^{i\theta} At\right] C \exp\left[e^{i\theta} Bt\right] dt$$

is the unique solution of

$$-AX + XB = C .$$

REMARK. If all $\alpha_{ij} < 0$, we may take $\theta = 0$ and

$$X = -\int_0^\infty e^{-At} Ce^{Bt}\, dt$$

is the unique solution of

$$-AX + XB = C.$$